

Anne-Marie Pflugbeil

**Stabilitätsanalyse des Membranproteins
HP0565-Eine experimentelle und
bioinformatische Studie**

BACHELORARBEIT

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Fakultät
Mathematik/Naturwissenschaften/Informatik

Mittweida, August 2010

Anne-Marie Pflugbeil

**Stabilitätsanalyse des Membranproteins
HP0565-Eine experimentelle und
bioinformatische Studie**

eingereicht als

BACHELORARBEIT

an der

HOCHSCHULE MITTWEIDA

UNIVERSITY OF APPLIED SCIENCES

Fakultät

Mathematik/Naturwissenschaften/Informatik

Mittweida, August 2010

Erstprüfer: Prof. Dr. rer. nat. Dirk Labudde
Zweitprüfer: Dipl.-Inf. (FH) Daniel Stockmann

Bibliographische Beschreibung:

Anne-Marie Pflugbeil:

Stabilitätsanalyse des Membranproteins HP0565-Eine experimentelle und bioinformatische Studie 2010. - 68 S. Mittweida,
Hochschule Mittweida - University of Applied Sciences,
Fakultät Mathematik/Naturwissenschaften/Informatik, Bachelorarbeit,
2010

Referat:

Diese Arbeit beschäftigt sich mit der Auswertung von Kraft-Abstandskurven aus einem SMFS-Experiment des Membranproteins HP0565. Gleichmaßen wurden die Möglichkeiten einer bioinformatischen Studie für eine Charakterisierung des Proteins genutzt. Im Hinblick auf die Möglichkeiten der Single-Molecule-Force Spectroscopy (SMFS) wurden Stabilitätsanalysen und statistische Betrachtungen durchgeführt.

Danksagung

An dieser Stelle möchte ich mich bei allen Personen bedanken, die mich bei der Erstellung dieser Arbeit unterstützt haben.

Mein Dank geht in erster Linie an meinen Betreuer Prof. Dr. rer. nat. Dirk Labudde für die tatkräftige Unterstützung, gute Zusammenarbeit, unzähligen Ideen und Engelsgeduld während der Erstellung dieser Arbeit. Ohne die zahlreichen Motivationsschübe hätten viele Grundgedanken für die Auswertung der Daten nicht umgesetzt werden können. Weiterhin möchte ich mich bei Dipl. Inf. (FH) Daniel Stockmann für die mühevollen informatischen Aufarbeitung der Daten, bedanken. Ein herzliches Dankeschön geht an Dr. Frank Dressel vom BIOTEC Dresden für die Hilfestellung und der Beantwortung offener Fragen bei Softwareproblemen. Zudem möchte ich Priv.-Doz. Dr. med. habil. Jörg-Olaf Habeck für die mikroskopischen Aufnahmen histologisch relevanter Präparate und mentale Unterstützung danken.

Florian Heinke danke ich für die Erstellung und Bereitstellung des Energieprofils. In diesem Zusammenhang möchte ich mich von Herzen bei der gesamten Projektgruppe für ein durchgehend angenehmes Arbeitsklima bedanken, bei dem der Spass nicht verloren ging und gegenseitige Unterstützung nie auf der Strecke blieb. Nicht zuletzt möchte ich mich von Herzen bei meinen Eltern und engsten Freunden bedanken, die in der Lage waren ein jedes Tief durch helfende Hände in ein Hoch zu verwandeln.

Mittweida, den 24.08.2010

Anne-Marie Pflugbeil

Inhaltsverzeichnis

Abbildungsverzeichnis	IV
Tabellenverzeichnis	V
Abkürzungsverzeichnis	VI
1 Einleitung	1
2 Zielstellung	3
3 Fachliches Umfeld	4
3.1 Membranproteine	4
3.2 Die Spezies <i>Helicobacter pylori</i>	5
3.2.1 Prävalenz und Verbreitung	5
3.2.2 Pathomechanismus des Bakteriums	5
3.3 Single-Molecule-Force-Spectroscopy	7
3.3.1 Prozedur	7
3.3.2 Das Kraftmikroskop als Sensor für molekulare Kräfte	8
3.3.3 Das Modell der wurmartigen Kette -Worm-like-chain-model-	11
3.4 Die Grenzen einer experimentellen Strukturbestimmung	12
4 Bioinformatische Charakterisierung von HP0565	13
4.1 Biologische Datenbanken	14
4.2 Charakterisierung der Primärstruktur	15
4.3 Sequenzähnlichkeitssuche mit Hilfe des „Basic Local Alignment Search Tool“	18
4.4 Multisequenzalignment ausgewählter Sequenzen	20
4.5 Sequenzbasierte Motivsuche	22
4.6 Beschreibung der Phylogenie	24
4.7 Vorhersage der 2D-Topologie	26
4.7.1 Grundlagen	26
4.7.2 Analyse	26
4.8 Vorhersage der 3D-Struktur von HP0565	30
4.8.1 Grundlagen	30
4.8.2 Analyse	31
5 Datenanalyse	35
5.1 Betrachtungen zum Experiment und experimentelle Eckdaten	35

5.2	Anwendung verschiedener Strategien für die Auswertung der FD-Kurven von HP0565-Datenfilterung-	39
5.3	Statistische Analyse vorhandener Entfaltungssignale	46
5.3.1	Verteilung möglicher Entfaltungssignale in Bezug auf den N-terminalen Datensatz	46
5.3.2	Verteilung möglicher Entfaltungssignale in Bezug auf den C-terminalen Datensatz	49
5.3.3	Verteilung von Kraft und Konturlänge in Bezug auf den N-terminalen Datensatz	51
5.3.4	Verteilung von Kraft und Konturlänge in Bezug auf den C-terminalen Datensatz	54
5.4	Spezifizierung von Stabilitätsbarrieren (Stabilitätsanalyse)	55
6	Ausblick	68
	Literaturverzeichnis	69
	Selbständigkeitserklärung	73

Abbildungsverzeichnis

3.1	Einteilung der Membranproteine	4
3.2	Schematische Darstellung eines Einzelmolekülkraftspektrometrischen Versuchs	8
3.3	Vereinfachte Kraftkurve	9
3.4	Typische Kraftkurve aus einem SMFS-Experiment	10
4.1	Analyseweg von HP0565	14
4.2	Sequenz HP0565 im FASTA-Format	17
4.3	MSA	21
4.4	Consensussequenz	22
4.5	Phylogramm	25
4.6	Kladogramm	25
4.7	Topologie von HP0565	28
5.1	Solubilisierung von integralen Membranproteinen	36
5.2	Kontrolliertes Entfalten von HP0565	38
5.3	Topologie mit gekennzeichneten Endpunkten möglicher Entfaltungen . .	41
5.4	FD-Kurve der N-terminalen Gruppe mit gekennzeichneten main- und side-Peaks	44
5.5	Kurvenalignment	45
5.6	Verteilung möglicher Entfaltungsevents im N-terminalen Datensatz	47
5.7	Konturlängenverteilung im N-terminalen Datensatz	49
5.8	Verteilung möglicher Entfaltungsevents im C-terminalen Datensatz	50
5.9	Konturlängenverteilung im C-terminalen Datensatz	51
5.10	Verteilung von Kraft und Konturlänge im N-terminalen Datensatz	52
5.11	Verteilung von Kraft und Konturlänge im C-terminalen Datensatz	55
5.12	Mögliche Entfaltungswege für den N-terminalen Datensatz	57
5.13	Mögliche Entfaltungswege für den N-terminalen Datensatz anhand der Topologie von HP0565	58
5.14	Mögliche Entfaltungswege für den C-terminalen Datensatz	60
5.15	Mögliche Entfaltungswege für den C-terminalen Datensatz anhand der Topologie von HP0565	61
5.16	Assoziation von Topologie, Sequenz und Entfaltungsbarrieren von HP0565 (N-terminal)	63
5.17	Assoziation von Topologie, Sequenz und Entfaltungsbarrieren von HP0565 (C-terminal)	64
5.18	Motivsuche in vorhergesagten transmembranen Regionen von HP0565 . .	67

Tabellenverzeichnis

4.1	Vorausgesagte Domänen	16
4.2	Datensatz der identifizierten ähnlichen Sequenzen zu HP0565	19
4.3	Ergebnis der Profilsuche bei Pfam	23
4.4	Topologische Bereiche	29
4.5	Annotierte Proteine des Stammes 26695	33
5.1	Mögliche Entfaltungsfälle	40
5.2	Signale entsprechend der 2D-Topology von HP0565	42

Abkürzungsverzeichnis

TM	Transmembran
SMFS	Single-Molecule-Force Spectroscopy
FD-Kurve	Kraft-Abstands-Kurve
MSA	Multisequenzalignment
As	Aminosäure (n)
aa	aminoacid (s)
HMM	Hidden-Markov-Modell
H.	Helix

Kapitel 1

Einleitung

Membranproteine spielen in der Strukturbiologie eine besondere Rolle. Die Zelle, als elementare Einheit, ist der Grundbaustein von Lebewesen. Durch Biomembranen werden die Zellen nach außen abgegrenzt und im Zellinneren werden unterschiedliche, voneinander getrennte Reaktionsräume geschaffen. Diese Kompartimente erlauben durch ihre steuerbare Durchlässigkeit eine selektive Permeabilität. Die charakteristische Kommunikation zwischen spezifischen Zellkompartimenten erfolgt sowohl über periphere als auch über integrale Proteine und Proteinkomplexe. Die wesentliche Bedeutung von Membranproteinen lässt sich besonders prägnant aus der Sicht der Genetik und Pharmazie verdeutlichen. Annähernd ein Drittel der uns bekannten Gensequenzen kodieren Membranproteine, welche wiederum mehr als die Hälfte aller existierenden Angriffspunkte für neuartige Therapeutika darstellen [EM00, MA02]. Somit ist das Wissen über die Struktur-Funktionsbeziehung dieser Proteine von zentraler Bedeutung für die Medizin und Pharmakologie. Hierfür ist eine Aufklärung der räumlichen Struktur dieser meist sehr großen Biomoleküle bis hin zu molekularen Details erforderlich. Die Single Molecule Force Spectroscopy (SMFS) ist eine junge Methode, um Membranproteine zu untersuchen. Gerade die inter- und intramolekularen Wechselwirkungen die ein Membranprotein in seiner natürlichen Faltung beschreiben, sind von großem Interesse und in den Fokus der Forschung gerückt. Beim Membranprotein HP0565 handelt sich um ein Protein von *Helicobacter pylori*. Das Gram negative Bakterium ist in der Lage die Magenschleimhaut des Menschen zu besiedeln und unter bestimmten Umweltbedingungen chronische Magenschleimhautentzündungen bis hin zu bösartigen Magengeschwüren herbei zu führen. Man geht davon aus, dass insbesondere Membranproteine, mit der Stellung als sogenannte Virulenzfaktoren, dafür verantwortlich sind. Das Membranprotein HP0565 der Spezies *Helicobacter pylori* wird in der Literatur als hypothetisch beschrieben, d.h. es liegen wenig bis keine Informationen über die die Funktion und molekulare Struktur des Proteins vor. Umso wichtiger ist eine bioinformatische Beschreibung bis hin zur Erstellung einer

3D-Struktur. Am Biotec der TU Dresden wurden SMFS-Experimente an diesem Protein durchgeführt. Zur Bestimmung von stabilisierenden Bereichen im Inneren des Membranproteins liegen mehrere hundert Messkurven für eine Interpretation dieser vor. Jedoch ist die Auswertung und anschließende Charakterisierung von unbekannten Membranproteinen noch nicht automatisiert. Durch die Einbindung von bioinformatischen Tools kann dieser Prozess beschleunigt und verallgemeinert werden. Aus heutiger Sicht ist die Modellierung von 2D-Strukturen und 3D-Strukturen ein fester Bestandteil der Bioinformatik. Wobei sich im Laufe der Bearbeitung des Projektes die Grenzen der uns zur Verfügung stehenden Möglichkeiten, die eine bioinformatische Analyse ausmachen, deutlich wurden.

Kapitel 2

Zielstellung

Kern der vorgelegten Arbeit stellte die Auswertung der vorliegenden Kraft-Abstandskurven über eine bereitgestellte Software [nan10] dar. Im Hinblick auf die Möglichkeiten der Single-Molecule-Force Spectroscopy (SMFS) wurden Stabilitätsanalysen und statistische Betrachtungen angestrebt. Diese Methode erlaubt es bestimmte Fragmente oder eben Proteine aus einer spezifischen Umgebung, die der Membran entspricht, herauszulösen. Durch das aufeinanderfolgende Lösen des Proteins aus der Membran heraus, entstehen einzigartige Entfaltungsmuster, die es erlauben Entfaltungswege spezifischer Proteine im Zusammenhang mit der aufgewendeten Kraft darzustellen. Für die Analyse von membranstabilisierenden Regionen sollte es zunächst möglich sein, den Datensatz auf charakteristische Entfaltungsmuster auftretender Sekundärstrukturelemente, zu untersuchen. Die im Voraus angestellte bioinformatische Charakterisierung von HP0565 verfolgte das Ziel, Informationen auf Sequenz-und Strukturebene für das hypothetische Membranprotein HP0565 zu erhalten und sinnvoll zu verknüpfen. Durch Informationen aus dieser bioinformatischen Beschreibung sollte es möglich sein, eine exakte 2D-Topologie und anschließend ein mögliches 3D-Modell vorherzusagen.

Kapitel 3

Fachliches Umfeld

3.1 Membranproteine

Zellmembranen bestehen aus einer Lipiddoppelschicht mit eingelagerten oder an dieser Schicht verankerten Proteinen. Die Membran grenzt eine Zelle nach außen hin ab und umgibt diese vollständig, so dass ein geschlossener Reaktionsraum vorliegt. Membranproteine können sich durch andere Proteine oder Lipide auf der Membran verankern oder sich innerhalb der Membran befinden. Demzufolge werden diese Proteine als Transmembranproteine bezeichnet (Abbildung 3.1) [Cle06].

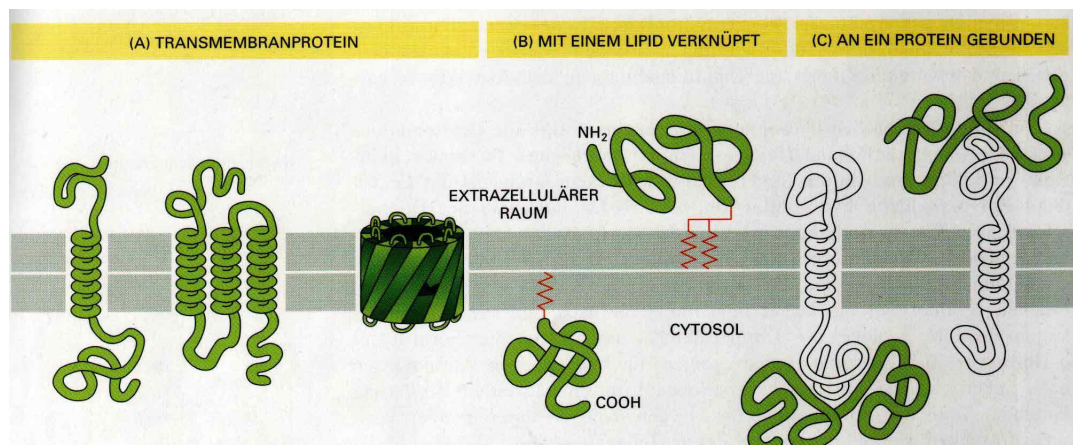


Abbildung 3.1: Einteilung der Membranproteine [Cle06]

Transmembranproteine durchqueren die Lipiddoppelschicht in Form einer alpha-Helix Domäne, wobei das polare Peptidgerüst nach innen zeigt und die hydrophobe Seite zum extrazellulären Raum ausgerichtet ist [Cle06]. Transmembranproteine werden je nach Anzahl der in der Membran integrierten Domänen in Singlepass-Transmembranproteine, welche die Membran nur einmal durchspannen und in Multipass-Transmembranproteine, die wiederum mehrmals die Membran durchspannen, unterteilt.

3.2 Die Spezies *Helicobacter pylori*

3.2.1 Prävalenz und Verbreitung

Den Studien zufolge sind aller Wahrscheinlichkeit nach mehr als die Hälfte der Weltbevölkerung mit *H.pylori* infiziert. Das Gram negative Bakterium ist mitunter in der Lage mehrere spezifische Enzyme, wie Katalasen, Ureasen, Oxidasen sowie mehrere Proteasen zu sezernieren, die eine Virulenz beachtlich erhöhen [Thi04]. Auch Membranproteine werden bei entsprechender Disposition in diesem Zusammenhang zu den Virulenzfaktoren gezählt. Vieles spricht dafür, dass die Infektion schon im Kindesalter innerhalb der Familie erworben wird, teilweise verbunden mit einer reduzierten basalen Magensäuresekretion. Die Übertragung der Infektion erfolgt vermutlich über den fäkal-oralen Weg durch kontaminierte Faeces oder auch gastrisch-oral durch erregerhaltigen Mageninhalt [Thi04].

3.2.2 Pathomechanismus des Bakteriums

Infektionen mit *Helicobacter pylori* verlaufen in der Regel chronisch. Obwohl chronische Magenschleimhautentzündungen fast immer ein Indiz für das Vorhandensein der Bakterien sind, bleiben jedoch die meisten *H.pylori*-positiven Individuen asymptomatisch. Zahlreiche Studien verdeutlichen gleichwohl den Zusammenhang zwischen einer Magenschleimhautbesiedlung durch *H.pylori* und Gastritis, peptischem Ulkus sowie die Bildung eines Magenkarzinoms [Gal05]. Typische Kolonisationsfaktoren von *H.pylori*, die ein andauerndes Verharren im Magenmilieu ermöglichen, sind neben der Expression spezifischer Enzyme und Proteine die Spiralform, die Beweglichkeit und die Fähigkeit zur Adhärenz des Bakteriums. Ein wesentlicher Auslöser für die gastroduodenale Ulkuserkrankung ist das Exprimieren des entzündungsfördernden vakuolisierenden Zytotoxins (Vac A Genprodukt), welches die Ausbildung von Vakuolen in Epithelzellen bewirkt. Diese füllen sich folglich mit Säure bis zum Zerplatzen, wodurch das umliegende Gewebe zerstört wird. Einem weiteren Mechanismus zufolge injiziert das Bakterium über einen nadelartigen Fortsatz ein Peptidoglycan in das Innere der Magenepithelzelle [Gmb].

Dieses Makromolekül dockt an einen Rezeptor an, wodurch eine Reaktionskette in Gang gesetzt wird, die schließlich zur Entzündung der Magenschleimhaut führt. Der Abschnitt auf dem Chromosom, der diesen Vorgang genetisch kodiert, wird als „Zytotoxin-assoziierte Gene-Pathogenitätsinsel“ bezeichnet [Gmb].

3.3 Single-Molecule-Force-Spectroscopy

Im folgenden Abschnitt sollen die Möglichkeiten der genannten Methodik und der theoretische Hintergrund erläutert werden. Die Single-Molecule-Force-Spectroscopy (SMFS) ist eine sehr junge Methode, die eine nähere Untersuchung, insbesondere von Membranproteinen, möglich macht und Einsichten in die Struktur-Funktionsbeziehungen der Biomoleküle erlaubt. Mittels SMFS wird es u.a. möglich eine Proteinstabilität auf molekularer Ebene und den Einfluss von Punktmutationen auf die Proteinstabilität zu betrachten. Weiterhin können gezielt Bindungsorte von Wirkstoffen an Targets bestimmt und Protein-Protein Wechselwirkungen betrachtet werden. Schon allein durch die genannten Aspekte wird das Potential dieser relativ neuen und innovativen Methodik deutlich. Dennoch existieren zum gegenwärtigen Entwicklungsstand eine Reihe von ungelösten Fragen und Problemen. Bisher wurden mit der Methode der Einzelmolekülspektroskopie nur wenige Membranproteine auf eine Struktur-Funktionsbeziehung untersucht. Der hohe Aufwand bei einer konventionellen Strukturaufklärung bedingt heutzutage die Entwicklung neuer Methoden, die es ermöglichen weitere Informationen über das Protein zu gewinnen.

3.3.1 Prozedur

An diesem Punkt soll kurz der Ablauf der Methode und wichtige Parameter konkretisiert werden, bevor zu einem späteren Zeitpunkt das Verfahren anhand des Datensatzes des zu analysierenden Proteins spezifiziert wird. Das Verfahren der SMFS lässt sich durch mehrere, aufeinanderfolgende Schritte erklären. Das Membranprotein befindet sich in einer Umgebung, die der natürlichen Membran entspricht oder ähnlich ist. Diese Oberfläche muss zunächst lokalisiert und abgebildet werden. Durch das Heranfahren der sogenannten Tastspitze an die Oberfläche der Membran, wird die Oberfläche auf die Proteine gescannt. Zwischen beiden Oberflächen wirken unspezifische Kräfte. Das Biomolekül haftet an die Tastspitze und wird somit aus der Membran gelöst und entfaltet. Die Entfaltungsevents werden in den resultierenden Kraft-Abstandskurven, entsprechend der aufgewandten Kraft und dem Lösen charakteristischer Sekundärstrukturelemente, dargestellt. Die FD-Kurve stellt eine Art Fingerprint dar, der für jedes Protein einzigartig ist [BCF08]. Dieser erlaubt molekulare Interaktionen, die die Stabilität vorhandener Sekundärstrukturelemente beschreibt, zu lokalisieren. Eine entsprechende Software [nan10] sorgt nun dafür, dass Kurven, die keine signifikante Entfaltung zeigen, extrahiert werden. Anschließend werden Ähnlichkeitsmaße für das Alignen der Kurven festgelegt. Für eine genaue Analyse der FD-Kurven durch den Betrachter, entsprechend der bestehenden Fragestellung, spielen Informationen aus der 3D-Struktur des Proteins eine wichtige Rolle.

3.3.2 Das Kraftmikroskop als Sensor für molekulare Kräfte

Das in den 1980er Jahren entworfene Atomkraftmikroskop wurde in den letzten Jahren zum Einzelmolekülkraftspektrometer (Abbildung 3.2) weiterentwickelt. Mit diesem kann die Kraft gemessen werden, die auf ein Protein einwirken muss, um es mechanisch zu entfalten [JNe00]. Die Kraftmikroskopie hat sich wie schon erwähnt, zu einer Methode entwickelt, mit der sich molekulare Kräfte an Grenzflächen im Detail analysieren lassen.

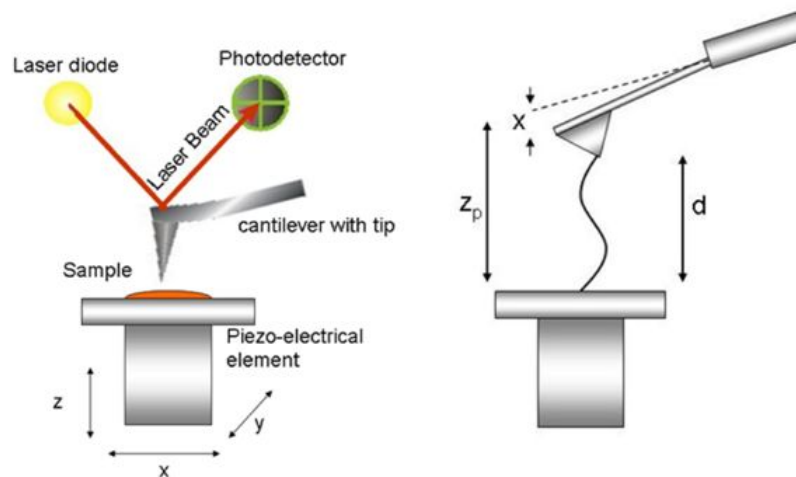


Abbildung 3.2: Schematische Darstellung eines Einzelmolekülkraftspektrometrischen Versuchs [LMS08]

Die Tastfeder (engl. cantilever) fungiert als Sensor, der auf die lokale Wechselwirkung zwischen Messspitze und Probe reagiert. Indem die Tastspitze an die Oberfläche heran und letztlich wieder zurückgefahren wird, kann eine Federauslenkung detektiert (Kraft-Abstanskurve) und in Abhängigkeit der Kraft von der Verschiebung des Piezostellglied, aufgezeichnet werden. Die Kraftauflösung liegt heute im Bereich von Piconewton und ist durch das thermische Rauschen an der Tastfeder eingegrenzt. Die Zeitspanne des gesamten Vorgangs liegt üblich im Sekundenbereich. Die Rückspur der Messkurven kann in einigen Fällen eine umfassende Form annehmen, die Rückschlüsse auf die Wechselwirkung zwischen Probe und Spitze zulässt. Während der Annäherung zur Probe ist es ebenso möglich, Informationen über die Oberflächenkräfte in räumlich begrenzten Bereichen zu sammeln. Die angesprochenen Wechselwirkungen umfassen u.a Van-der-Waals-Wechselwirkungen im nN-Bereich sowie entropische Kräfte von wenigen 100pN. Diese treten unter anderem beim Strecken von Einzelmolekülen auf. Die Art der Wechselwirkungen ist stark von den experimentellen Bedingungen abhängig. So beeinflussen das Probenmaterial und chemische Modifizierungen der Oberfläche das Auftreten bestimmter Kräfte [JNe00].

Die Auslenkung der Tastfeder ist ein Maß für die Wechselwirkung zwischen Tastspitze und Probe. Wirken attraktive Kräfte (Kraft < 0) auf die Tastfeder, so wird diese in negative Richtung abgelenkt. Bei einer repulsiven Kraft erfolgt analog eine positive Auslenkung der Feder. Für die Annäherung und Rückführung der Tastfeder von bzw. zur Probe lassen sich signifikante Kurvenverläufe feststellen. Befindet sich die Tastnadel in großer Entfernung zur Probe bleibt die Position der Nadel unverändert (1) (Abbildung 3.3). Bei einer Annäherung wirken attraktive Oberflächenkräfte bis das Stabilitätskriterium nicht mehr erfüllt ist. In diesem Fall spricht man von unspezifischen Wechselwirkungen (2). Die Tastnadel tritt nun abrupt in repulsiven Kontakt (2)-(3) und deformiert somit die Probenoberfläche. Die resultierende Kraft ist dementsprechend negativ, obwohl auch abstoßende Kräfte wirken. Die resultierende Verbiegung der Tastfeder ist vorwiegend von der Viskosität der Unterlage abhängig (3)-(4).

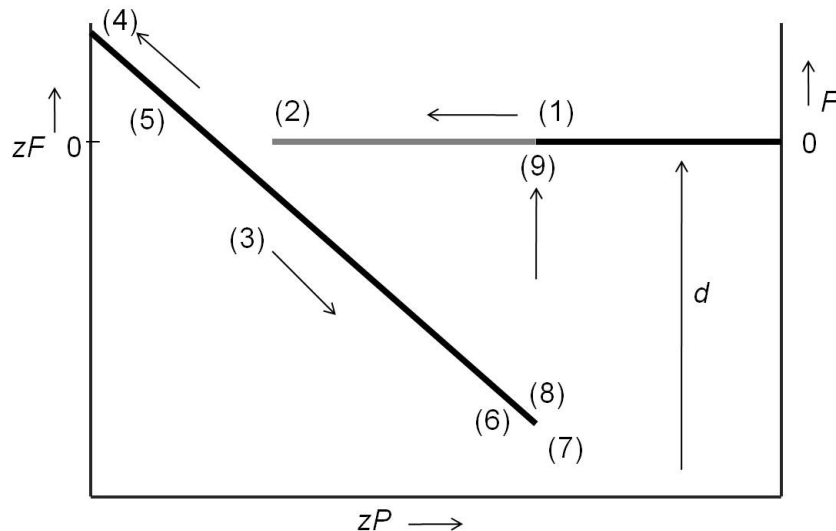


Abbildung 3.3: Vereinfachte Kraftkurve (modifiziert nach [JNe00]). Die vereinfachte Darstellung einer Kraftkurve zeigt das Heranführen der Tastfeder bis zum Probenkontakt (1-4) und das anschließende Entfernen bis zum Abreißen der Probe (5-9). Die Federauslenkung wird mit dem Parameter zF und die Probenverschiebung mit dem Parameter zP festgelegt. d ist der Abstand zwischen Cantilever und Probe.

Die Anwendung des Kraftmikroskops reicht an dieser Stelle von Eindrucksmessungen an Elastomeren oder dünnen organischen Schichten bis hin zu mechanischen Experimenten an einzelnen Zellen. Während der Rückspur der Tastnadel nehmen die repulsiven Kräfte wieder allmählich ab (5)-(6) bis eine Trennung der Nadel von der Probenoberfläche erfolgt (6). Das Minimum der Kurve (7) stellt die maximal messbare, attraktive Kraft bei dem Vorgang dar. Sie entspricht der Adhäsionskraft. Analog zur unspezifischen Wechselwirkung liegt auch hier eine mechanische Instabilität vor. Der Kontakt von Tastfederspitze und Probe wird demnach auch hier schlagartig getrennt (8)-(9) [JNe00]. Da die Versuchsanordnung nahe des Minimums (7) sehr empfindlich auf Störungen reagiert, wie z.B auf thermisches Rauschen oder Gebäudeschwingungen, findet die Trennung der Tastspitze schon beim Kurvenminimum statt. Eine Kraftkurve beschreibt die gemessene Kraft aufgetragen gegen die Verschiebung des z -Piezostellglieds. Die angesprochene Verschiebung lässt sich in einen Spitze-Probe-Abstand umrechnen, indem von der Federauslenkung die Kontaktgerade subtrahiert wird. Entsprechend dieser Normierung erhält man eine Kraft-Abstandskurve [JNe00] (Abbildung 3.4).

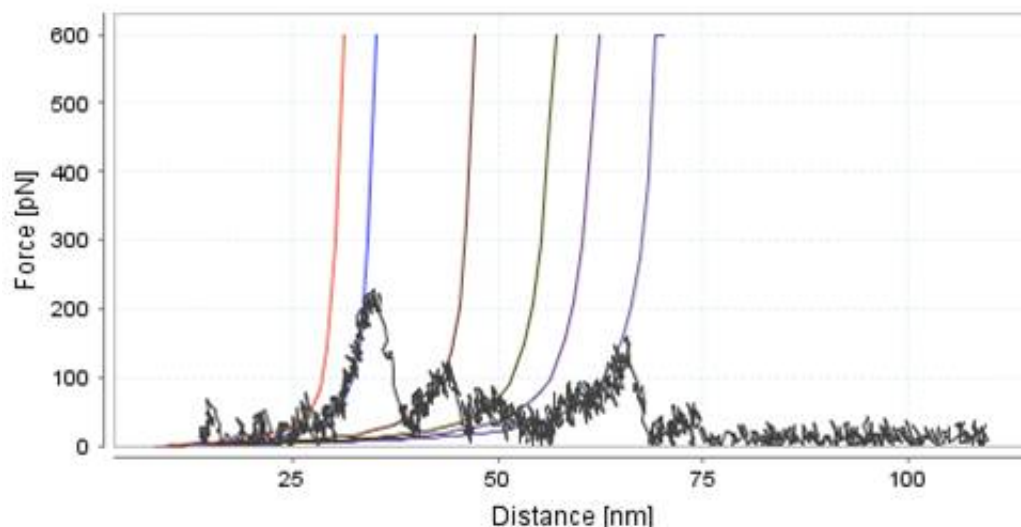


Abbildung 3.4: Typische Kraftkurve aus einem SMFS-Experiment [nan10]. Abgebildet ist eine resultierende Kraft-Abstandskurve von HP0565 mit WLC-Fit (Siehe Kapitel Das Modell der wurmartigen Kette -Worm-like-chain-model-)

3.3.3 Das Modell der wurmartigen Kette

-Worm-like-chain-model-

Wird ein Polymer als ein homogener Faden mit konstanter Biegefestigkeit betrachtet, führt dies zum Modell der sogenannten wurmartigen Kette, das aufgrund der Konformation des Moleküls durch eine kontinuierliche Raumkurve definiert wird [JNe00]. Der mathematische Ansatz, der das Modell beschreibt, kennzeichnet die Flexibilität des Moleküls mit der Persistenzlänge. Dieser Parameter trifft Aussagen über die Faltungsneigung des Proteins und ist stark abhängig von der aufgewendeten Kraft. Mit anderen Worten: Je größer die Persistenzlänge, desto höher ist die Tendenz, dass sich das Molekül knäuelte [JNe00]. In Bezug auf die Bewertung von FD-Kurven lassen sich durch das Modell Aussagen über die Polypeptidlängen entfalteter Elemente treffen. Diese Länge entspricht der aus dem Modell resultierenden Konturlänge. Bricht die Tertiärstruktur an einer Stelle unter Krafteinwirkung auf, so wächst die Konturlänge des Polypeptids an. Mit dem Konturlängen-Parameter werden folgend Stabilitätsbarrieren gekennzeichnet. Durch eine statistische Analyse von ermittelten Konturlängen, können spezifische Entfaltungswege des Proteins untersucht werden. Hierbei findet eine Unterscheidung zwischen Haupt- und Nebenentfaltungswegen statt. Als Maß für diese Einteilung wird dabei das quantitative Auftreten vorhandener Konturlängen gesehen.

3.4 Die Grenzen einer experimentellen Strukturbestimmung

Betrachtet man den experimentellen Weg einer Strukturbestimmung von Biomolekülen, so wird heutzutage in der Praxis auf zwei umfassende Methoden der Strukturaufklärung zurückgegriffen. Zum Einen die Kristallografie (engl. X-ray) und zum Anderen die Nuclear-Magnetic-Resonanzspektroskopie (NMR). Allerdings weisen beide Verfahren deutliche Grenzen bei der Strukturaufklärung von Membranproteinen auf. Für eine Strukturbestimmung über NMR sind Membranproteine, strukturell betrachtet, zu komplex und eine Kristallisation ist oftmals nur unter erschwerten Bedingungen möglich. Diese Problematik lässt sich wie folgt erklären: Bei einem Membranprotein lassen sich topologisch drei Domänen abgrenzen. Von denen entsprechen zwei den extramembranösen hydrophilen Bereichen des Proteins auf beiden Seiten der Membran. Hier besteht Kontakt zur wässrigen Phase oder mit den polaren Kopfgruppen der Lipide. Die hydrophobe Domäne durchspannt die Lipiddoppelschichtmembran. Innerhalb der Membran besteht Kontakt zu den Alkylketten der Lipide. Durch die unterschiedlichen Domänen sind solche Proteine nach dem Herauslösen aus der Membran sowohl in polaren als auch in apolaren Medien unlöslich [Wer06]. Man behilft sich durch den Einsatz von Detergentien, die den hydrophoben Bereich bedecken und dem Protein auch dort eine hydrophile Oberfläche verleihen. Durch Kontaktaufnahme mit der Universität Konstanz und dem FMP Berlin, erhielten wir im Verlauf der Bearbeitung des Projektes den Hinweis, dass das „New York Consortium on Membrane Protein Structure“ seit geraumer Zeit versucht, dieses Protein zu kristallisieren. Jedoch ohne Erfolg. Das lässt den Schluss zu, dass zum gegenwärtigen Zeitpunkt eine Strukturaufklärung von HP0565 über X-Ray ausgeschlossen ist. Auf der experimentellen Ebene können deshalb zusätzliche Informationen über das Prinzip der SMFS erlangt werden.

Kapitel 4

Bioinformatische Charakterisierung von HP0565

Die angestellte bioinformatische Charakterisierung von HP0565 verfolgte das Ziel, Informationen auf Sequenz- und Strukturebene für das hypothetische Protein HP0565 zu erhalten und sinnvoll zu verknüpfen. Durch ein möglichst breites Spektrum an Einzelinformationen, die mit Hilfe bioinformatischer Werkzeuge und Datenpools ermittelt wurden, sollten Korrelationen zwischen den einzelnen Ebenen einer Sequenz- und Strukturanalyse (Abbildung 4.1) geschaffen werden. Unter Einbindung dieser Informationen, konnten ebenfalls evolutionäre Aussagen getroffen und phylogenetische Verwandtschaften untersucht werden. Der Zusammenhang von Sequenz, Struktur und Funktion bleibt untrennbar und bildet die Grundlage für die Analyse von bisher nicht beschriebenen Membranproteinen, wie HP0565 eines darstellt. Durch Informationen aus dieser bioinformatischen Beschreibung, sollte es möglich sein, eine exakte 2D-Topologie und anschließend ein mögliches 3D-Modell vorherzusagen. Im Hinblick auf die Analyse der vorhandenen FD-Kurven verfolgte die durchgeführte bioinformatische Beschreibung das Ziel, eine 3D-Struktur von HP0565 zu erstellen, die es im günstigsten Fall ermöglicht, vorhandene Entfaltungswege besser zu verstehen.

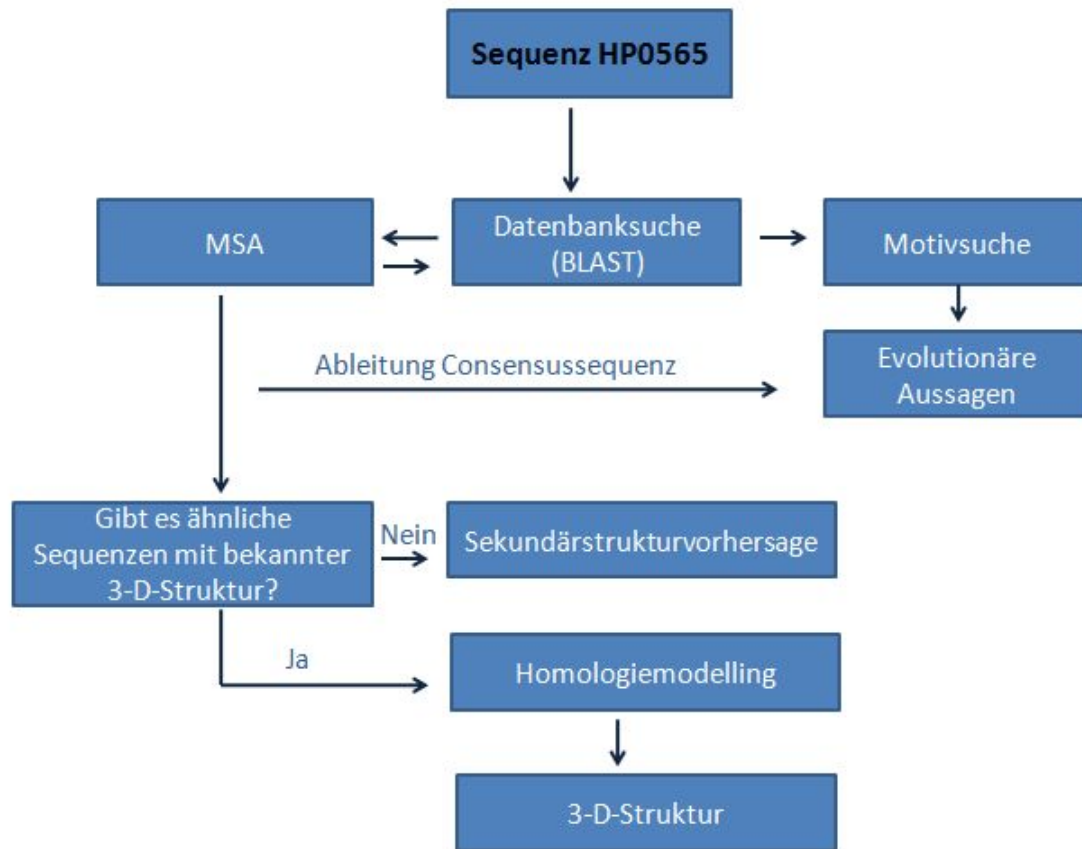


Abbildung 4.1: Analyseweg von HP0565. Anhand der Abbildung soll der grobe Analyseweg von der Sequenz, als Ausgangsinformation, über die Phylogenie bis hin zur Strukturanalyse aufgezeigt werden. Ausgangspunkt einer wissenschaftlichen bioinformatischen Analyse ist in der Regel die Sequenz des unbekannten Proteins. Auf sequenzieller Ebene schloss sich im ersten Schritt eine Datenbanksuche nach ähnlichen Sequenzen zu HP0565 an und in dem Zusammenhang ein Multisequenzalignment (MSA) mit diesen. Weiterführend war es möglich phylogenetische Verwandtschaftsbeziehungen zu untersuchen.

4.1 Biologische Datenbanken

Die für die Sequenzanalyse von HP0565 erforderlichen Daten wurden voranging aus den Datendomänen des „National Center for Biotechnology Information“ (NCBI) und dem „European Bioinformatics Institute“ (EBI) bezogen. Als zentrale Quelle für Informationen zu den Proteinen und Proteinfunktionen diente die primäre Datenbank „Universal Protein Resource-Knowledgebase“ (UniProt KB), welche der Datendomäne UniProt zugehörig ist. Weiterhin werden die Datenbanken Uniref und UniPark zu der Datendomäne UniProt gezählt. Kategorisch werden die Daten der UniProtKB in zwei große Gruppen eingeteilt, die unter der Bezeichnung SWISS-PROT und TrEMBL laufen [Gae07]. Bei

der Datenbank-Kategorie SWISS-PROT [BBAB02] handelt es sich um manuell ausgewertete und annotierte Daten, die neben der Sequenz noch zusätzliche Informationen wie Domänen, Motive und postranslationale Modifikationen [Gae07] umfassen. Der Datensatz der Datenbank, welcher aus einer computergestützten Auswertung von Proteinsequenzen hervorgeht wird mit TREMBL [BBAB02] bezeichnet, wobei hier die Einträge nicht voll manuell bearbeitet sind.

Als wichtigstes Werkzeug für eine Extraktion von ähnlichen Sequenzen zu HP0565 bewies sich das „Basic Local Alignment Search Tool“ (BLAST). Weiterhin wurde für den Versuch eine 3D-Struktur von HP0565 zu erstellen zunächst auf die Protein-Datenbank (PDB) [BWF99] zugegriffen, um Strukturen von möglichen Homologen aufzufinden. Die PDB speichert Daten aus experimentell abgeleiteten 3D-Strukturen von Proteinen [Gae07], aber auch von Nukleinsäuremolekülen. Im Gegensatz zur PDB verwaltet die Strukturdatenbank ModBase [PEB04] dreidimensionale errechnete Protein-Modelle, welche durch das vergleichende modellieren von Strukturen zustande kommen und sich nicht auf experimentelle Daten stützen. Weitere Server, die für eine Vorhersage einer 3D-Struktur Anwendung fanden, waren: $(PS)^2$ -Server [CCJKJM06] und I-TASSER. Im Gegensatz zum $(PS)^2$ -Server, dessen Methode auf einer homologiebasierten Vorhersage beruht, stützt sich I-TASSER auf ein Profile- Profile threading Alignment [Zha08].

4.2 Charakterisierung der Primärstruktur

Unter vorliegender Sequenz von HP0565, wurden zunächst Parameter wie die chemische Zusammensetzung und die Unterteilung in Domänen näher untersucht. Für alle angestrebten Betrachtungen, stand die experimentell ermittelte Sequenz zur Verfügung. Bei der Suche nach der Sequenz in Sequenzdatenbanken stellt der Betrachter fest, dass diese Sequenz ohne Regionen geringer Komplexität hinterlegt wurde. Meist sind solche Bereiche mit nacheinander befindlichen, oft wiederholenden Aminosäuren gekennzeichnet. Domänen werden in der Literatur als funktionelle Untereinheiten deklariert, die im Gesamten die Funktion des Proteins bestimmen. Damit wird auch das Ziel dieser Analyse klar. Können wir signifikante Domänen auf sequentieller Ebene ausmachen, so wäre es möglich, eventuell Aussagen über den Zusammenhang strukturelle Untereinheit-Funktion von HP0565 zu treffen. Für eine nähere Betrachtung einzelner Segmente der vorliegenden Sequenz, diente die Datenbank SMART. Die Analyse ergab vier transmembrane Bereiche und zwei Bereiche (Tabelle 4.1), die dennoch mit der Eigenschaft einer geringen Komplexität gekennzeichnet wurden. Auch wenn diesen Bereichen bisher noch keine große Bedeutung zugesprochen wird, zählen sie zu Motiven, die eine intrinsische Fehlordnung

[DCR05] innerhalb der Proteinsequenz zeigen. Basiert die Fehlordnung auf einer signal-gesteuerten Wechselwirkung, die hochspezifisch, jedoch nicht umkehrbar ist, können diese Bereiche eine Art Verbindung zu anderen Kompartimenten mit hoher Spezifität und geringer Affinität herstellen [DCR05].

Name	Beginn [As]	Ende [As]
transmembran	20	39
low complexity	75	86
transmembran	104	126
transmembran	128	156
transmembran	165	187
low complexity	212	229

Tabelle 4.1: Vorausgesagte Domänen

```
>hp0565|SEQUENCE
MGGSQALKSLLEVITKLQNLGGYLMHIAIFIIFIWIGLKFVPYEAEGIAPFVANSPPFS
FMYKFEKPAYKQHKMSSESQSMQEEMQDNPKIVENKEWHKENRTYLVAEALGITIMILGIL
VLLGLWMPLMGVIGLLVAGMTITTLSTLFTTPEVFNQHFPPWLSGAGRLVVKDLALFAG
GLFVAGFDAKRYLEGKGFCLMDRSSVGITKCSSGCCSAAAGSHHHHHH
```

Abbildung 4.2: Sequenz HP0565 im FASTA-Format. Hier dargestellt, die experimentell ermittelte Sequenz von HP0565 im FASTA-Format. Blau gekennzeichnete Abschnitte deuten die transmembranen Domänen an. Bereiche, welche grau gekennzeichnet wurden, zeigen die detektierten Regionen geringer Komplexität an.

Chemisch betrachtet tauchen die Aminosäuren Leucin (11.4%), Glycin (10,5%), Alanin (7,4%) mit einem prozentualen Anteil von insgesamt 29%, bezogen auf die absolute Aminosäurezusammensetzung des Proteins, am häufigsten auf [Ser].

4.3 Sequenzähnlichkeitssuche mit Hilfe des „Basic Local Alignment Search Tool“

Das Ziel einer Sequenzähnlichkeitssuche ist das Auffinden ähnlicher Sequenzen zur Anfragesequenz, um somit Informationen von einer bereits beschriebenen Sequenz auf eine bisher unbeschriebene zu übertragen. Ausgangspunkt der Ähnlichkeitssuche war die Anfrage mit der Sequenz von HP0565 im FASTA-Format. Im Focus lagen zunächst Ergebnisse, die über eine WU-BLAST2 Protein Suche ermittelt wurden, welche sich durch einen geringen Verlust an Sensitivität während der Suche, auszeichnet. Für nachfolgende Schritte wurden Sequenzen aus der UniProtKB Swiss-Prot bevorzugt verwendet. Die Bewertung des Sequenzalignments erfolgte über die Auswahl der Blosom62-Matrix. Aus der Ähnlichkeitssuche resultierten eine Vielzahl von paralogen Sequenzen zu HP0565, ausgehend von der Spezies *Helicobacter pylori*. Zwar wiesen diese eine bezeichnend hohe Identität zu HP0565 auf, jedoch sind die Sequenzen zum gegenwärtigen Zeitpunkt ebenso unbeschriebenen Charakters und dementsprechend begrenzt geeignet für anschließende Schritte der Analyse, insbesondere für einen Rückschluss auf die Struktur, da sich keine Informationen von diesen Proteinen auf HP0565 übertragen lassen. Demzufolge wurden bei der Auswahl Sequenzeinträge bevorzugt, die in einem hohen Maß beschrieben sind (Tabelle 4.2) bis hin zu vorhandenen strukturellen Präferenzen oder der Möglichkeit, dass der Quellorganismus aufgrund seines Stoffwechsels Parallelen zu *Helicobacter pylori* aufwies. Regionen geringer Komplexität blieben unmaskiert. Um die Suche nach ähnlichen Sequenzen zu verfeinern, wurde eine PSI-BLAST Suche angeschlossen. Jedoch ergaben sich dadurch keine neuen nützlichen Informationen.

Nr.	Datenbank-ID	Herkunft	Länge [As]	Identität [%]	Score	e-value
1	UNIPROT: Y219-HAEIN	<i>Haemophilus influenzae</i>	213	40	271	3.1e-22
2	UNIPROT: A7V5S9-BACUN	<i>Bacteroides uniformis</i>	196	38	271	3.1e-22
3	UNIPROT: Q5LFG0-BACFN	<i>Bacteroides fragilis</i>	200	35	271	3.1e-22
4	UNIPROT: Q64W83-BACFR	<i>Bacteroides fragilis</i>	200	35	271	3.1e-22
5	UNIPROT: YKGB-PASHA	<i>Pasteurella haemolytica</i>	211	40	265	1.3e-21
6	UNIPROT: A7JVV3-PASHA	<i>Mannheimia haemolytica</i>	212	40	265	1.3e-21
7	UNIPROT: Q8ZR55-SALTY	<i>Salmonella typhimurium</i>	186	34	235	2.0e-18
8	SW: YKGB-Ecoli	<i>Escherichia coli</i> (K12)	197	31	227	7.6e-19

Tabelle 4.2: Datensatz der identifizierten ähnlichen Sequenzen zu HP0565

Der in der Tabelle verzeichnete Satz an Sequenzen, zeigt das Ähnlichkeitsmaß dieser Sequenzen zu HP0565. In Bezug zur ersten aufgeführten Sequenz (Y219-HAEIN) ist zu sagen, dass jene prozentual die höchste Ähnlichkeit zur Abfragesequenz aufweist. Eine hohe Ähnlichkeit konnte auch durch Genomanalysen der beiden Spezies *Helicobacter pylori* und *Haemophilus influenzae* zueinander nachgewiesen werden. Dabei offenbarten sich Genprodukte, die für bestimmte pathogene Faktoren beider Organismen gleichermaßen verantwortlich sind. Jedoch ist dieses Protein ebenfalls hypothetisch. Die Auswahl der aufgeführten Sequenzen lieferte im Hinblick auf eine Strukturvorhersage von HP0565 erste Anhaltspunkte. Die vorgenommene Ähnlichkeitssuche weist nur wenige hilfreiche Informationen hinsichtlich einer quantitativen Erfassung von Ähnlichkeiten aller Sequenzen auf. Wollen wir also mehr über konservierte Abschnitte wissen, muss sich an diesem Punkt gezwungenermaßen ein sogenanntes Multisequenzalignement anschließen.

4.4 Multisequenzalignment ausgewählter Sequenzen

Als Multisequenzalignment wird eine Sequenzausrichtung von mindestens drei Sequenzen [Les02] bezeichnet, um schnell Übereinstimmungen zwischen diesen zu erkennen. Ziel soll nicht nur eine Erfassung von Ähnlichkeiten zwischen den Sequenzen sein, sondern auch das Feststellen von konservierten Domänen und Motiven [Les02]. Dadurch können aus dem MSA Rückschlüsse auf entwicklungsgeschichtliche Verwandtschaftsverhältnisse und auch auf Struktur gezogen werden. Für die Erstellung des MSA wurde das Programm ClustalW2 verwendet, welches auf dem Webserver des European Bioinformatics Institute beheimatet ist. ClustalW2 ist ein völlig automatisches Programm für die globale, progressive Ausrichtung von DNA und Protein-Folgen und bietet zusätzlich die Möglichkeit Verwandtschaftsbeziehungen durch die Erstellung von phylogenetischen Bäumen nachzuvollziehen [LBB07]. Das Programm berechnet das beste Match für die ausgewählten Folgen, so dass Identitäten, Ähnlichkeiten und Unterschiede sichtbar werden.

Bei Integration aller aus dem Datensatz erfassten neun Sequenzen waren 46 Aminosäuren der 229 Aminosäuren zu 100% konserviert (Abbildung 4.3). Diese Bereiche haben sich demnach während der Entwicklungsgeschichte in allen aneinander ausgerichteten Sequenzen nicht verändert. Wurden lediglich nur die Sequenzen Y219-HAEIN, YKGB-PASHA und YKGB-ECOLI mit der Anfragesequenz ausgerichtet, waren 51 Aminosäuren der rund 229 Aminosäuren vollkommen identisch. Auffällig war, dass Regionen geringer Komplexität fast ausschließlich in der Sequenz von HP0565 auftauchten. Aus diesem Grund war davon auszugehen, dass infolge evolutionärer Veränderungen diese Bereiche hinzugefügt wurden. An diesem Punkt sollte noch erwähnt sein, dass low-complexity-Regionen in der Sequenz von HP0565 für nachfolgende Motiv- und Strukturuntersuchungen ausgeschlossen wurden. Damit sollte falsch-positiven Vorhersagen aus dem Weg gegangen werden. Sucht der Interessierte in bekannten Sequenzdatenbanken nach HP0565, findet dieser stets eine Sequenz mit 215 Aminosäuren vor, ohne die Bereiche einer geringen Komplexität.



Abbildung 4.3: MSA [LBB07]. Grau hinterlegte Bereiche kennzeichnen 100% konservierte Aminosäuren, die in allen neun Sequenzen vorkommen. Türkise Bereiche verweisen auf Low-complexity-Regionen in HP0565.

Die „beste“ Sequenz, die aus einem Sequenzalignment mehrerer Sequenzen heraus entsteht wird als Consensussequenz bezeichnet (Abbildung 4.4). Für die Ermittlung der Consensussequenz des vorliegenden Multisequenzalignments wurden ausschließlich die ermittelten identischen Aminosäuren herangezogen.

```

M-----I---WIG-LK---YEA-I---FV---NSP---S---Y-----Y---E-----W---N-TY
---G---L-----G-----TLSF---TTPE---V-----FP-L---GRL---KD---MM-----L
    
```

Abbildung 4.4: Consensussequenz. Lücken zwischen identischen Aminosäuren repräsentieren ähnliche Aminosäuren der alignierten Sequenzen.

4.5 Sequenzbasierte Motivsuche

Als Sequenzmotive werden konservierte Regionen bezeichnet, die im günstigsten Fall mit der für das Biomolekül charakteristischen Funktion korrelieren. Da noch keine Hinweise auf die Funktion des hypothetischen Proteins HP0565 vorliegen, sollte die Suche nach eventuell konservierten Bereichen in der Sequenz neue Informationen liefern, um ebenfalls mögliche Verwandtschaftsbeziehungen zu bereits beschriebenen Proteinen ableiten zu können. In der Regel werden Motive aus der vom MSA abgeleiteten Consensussequenz formalisiert. Im Verlauf wurden zwei Motiv-Analysen durchgeführt. Die erste umfasste eine rein formale Mustersuche in der sequenziellen Abfolge der Aminosäuren von HP0565. Zu einem späteren Zeitpunkt war es möglich eine zweite Analyse durchzuführen, die im Kontext zur Auswertung und Charakterisierung vorhandener FD-Kurven, weiterführende strukturelle Präferenzen in Bezug auf die 2D-Topologie liefern sollte, jedoch in an dieser Stelle noch keine Rolle spielt. Die sequenzbasierte Betrachtung von Motiven erfolgte unter Zuhilfenahme der sekundären Datenbanken PROSITE und Pfam. Speziell Pfam konzentriert sich eher auf eine globale Sequenzähnlichkeit zwischen Proteinsequenzen [Les02]. Dabei wird mit einer positionspezifischen Matrix die Auftretenswahrscheinlichkeit bestimmter As an bestimmten Positionen festgelegt. Profile repräsentieren funktionell interessante Domänen. Die Anfrage mit der Sequenz von HP0565 auf funktionell-relevante Motive lieferte keine Ergebnisse. Auch die Suche nach Motiven in den übrigen Sequenzen aus dem MSA war wenig aufschlussreich. Die Motivdatenbank PROSITE ermöglicht dem Nutzer ein definiertes Muster, nach dem in der entsprechenden Sequenz gesucht werden soll, in der PROSITE-Notation anzugeben. Jedoch konnten keine Treffer für die aus der Consensussequenz abgeleiteten Muster erzielt werden. Eine Analyse über Pfam ergab einen signifikanten Treffer und einen insignifikanten Treffer (Tabelle 4.3). Als signifikanter Treffer wurde die DUF417 Familie angegeben. Eine Familie setzt

sich aus einer Sammlung von Pfam-A Einträgen zusammen [Les02], die durch Ähnlichkeiten in der Sequenz, Struktur oder HMM-Profilen beschrieben ist. Pfam-A Treffer sind weitaus signifikanter als Pfam-B Treffer, da diese genau untersuchte Profile aus multiplen Alignments umfassen.

Pfam-A	Beschreibung	Typ	Sequenzbereich [As]	HMM	e-value
DUF417	Familie uncharakterisierter Proteine	Familie	8 bis 193	1 bis 200	2.2e-125

Tabelle 4.3: Ergebnis der Profilsuche bei Pfam. Die Übersicht beinhaltet lediglich den signifikanten Treffer der Suche. Aufgrund einer zu geringen Signifikanz des 2 Treffers wurde dieser ausgeschlossen.

4.6 Beschreibung der Phylogenie

Im phylogenetischen Kontext ist ein Baum bzw. ein Stammbaum eine mathematische Konstruktion, welche die stammesgeschichtlichen Verwandtschaftsverhältnisse von einer bestimmten Gruppierung widerspiegelt [HL03]. Im Falle der molekulare Phylogenetik werden zur Konstruktion „molekularer Stammbäume“ Sequenzdaten verwendet, die vorzugsweise aus multiplen Alignments stammen. Für die Erstellung dieser Bäume berechnen distanzbasierte Methoden die Distanz für alle Sequenzpaare eines Alignments [KM08].

Die Darstellung der Sequenzen erfolgte im Falle der vorliegenden Analyse für das Membranprotein HP0565 in zwei unterschiedlichen metrischen Stammbaumtypen. Das Kladogramm, die wohl einfachste Form, spiegelt ausschließlich die Topologie bzw. den Verzweigungsgrad des Baumes wider. Im Gegensatz zum Kladogramm liefert ein Phylogramm zusätzliche Informationen über Veränderungen in den Astlängen, wie zum Beispiel die Anzahl an Substitutionen [HL03]. Als Instrument für die Erstellung des phylogenetischen Baumes wurde das Multisequenzalignment in Abbildung 4.3 genutzt. Bei Betrachtung und Beurteilung der phylogenetischen Bäume musste beachtet werden, dass BLAST zwar eine Reihe ähnlicher Sequenzen zu HP0565 lieferte, diese aber aus unterschiedlichen Organismen stammten. Folglich handelt es sich im eigentlichen Sinne weder um „echte“ Paraloge und aufgrund der zum Teil fehlenden Funktionen der verwendeten Sequenzen, um wirkliche Orthologe. Aus dem Phylogramm (Abbildung 4.5) lies sich jeweils leicht die Gruppierung der gefunden ähnlichen Sequenzen aus den Organismen *Bacteroides uniformis* und *Bacteroides fragilis* und die Gruppierung aus *Haemophilus influenzae*, *Pasteurella haemolytica* sowie *Mannheimia haemolytica* erkennen. Setzte man HP0565 als Wurzel, so war der evolutionäre Abstand zur erst genannten Gruppierung am größten. Wiederum war die Distanz zu *Haemophilus influenzae*, *Pasteurella haemolytica*, *Mannheimia haemolytica* sowie *Escherichia coli* am geringsten, was wiederum den Zusammenhang der Spezies *Helicobacter pylori* und *Haemophilus influenzae* und *Escherichia coli* beweist. Auffällig war, dass die Distanzen von YKGB-PASHA und A7JVV3-PASHA zu HP0565 weitaus geringer als die der Gruppe von BACFN, BACUN, BACFR zu HP0565 waren. Wurden jedoch die einzelnen Übereinstimmungen aus den Alignments der genannten Sequenzen zu HP0565 betrachtet, dann fiel auf, dass sich mehr Überstimmungen zwischen HP0565 und der Gruppierung von BACFN, BACUN und BACFR finden lassen. Eine reine Topologiedarstellung der einzelnen Gruppen lieferte das vorliegende Kladogramm (Abbildung 4.6).

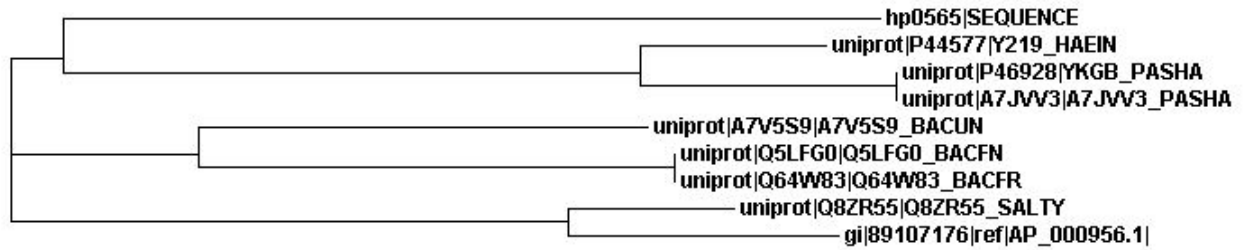


Abbildung 4.5: Phylogramm [LBB07]. Phylogramm der Sequenzen aus dem MSA, erstellt mittels ClustalW2. Als Wurzel wurde HP0565 ausgewählt.

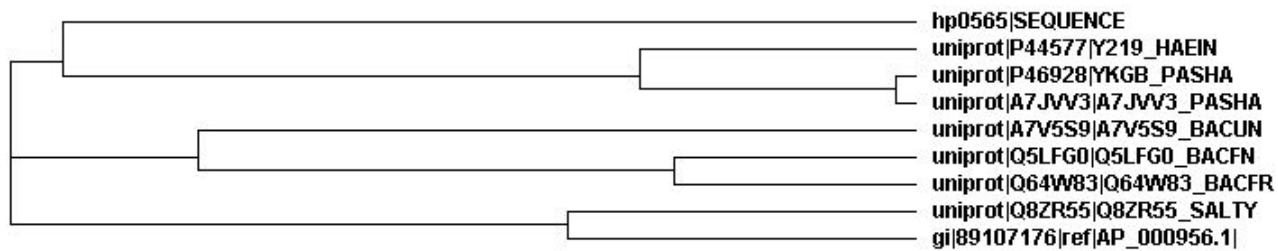


Abbildung 4.6: Kladogramm [LBB07]. Kladogramm für eine Topologiedarstellung.

4.7 Vorhersage der 2D-Topologie

4.7.1 Grundlagen

Die rechnergestützte Erkennung helikaler Membranabschnitte unter Zuhilfenahme der Aminosäuresequenz gilt als nützliches Tool in der Bioinformatik. Die hierfür bereitstehenden Vorhersageprogramme stützen sich größten Teils auf Algorithmen, die Eigenschaften und Nachbarschaften von Aminosäuren mit Hilfe von Neuronalen Netzen, dynamischer Programmierung bzw. Hidden-Markov Modellen bewerten. Eine Einarbeitung von multiplen Alignment Scores in einige Methoden beabsichtigt eine Verbesserung der Vorhersagekraft. Kontaktflächen zwischen Helices sind überwiegend konserviert und lassen sich deshalb oft von den ebenfalls hydrophoben Kontaktflächen zwischen den TM-Helices und der Membran unterscheiden [Hil03]. Die Abfolge zytosolischer und luminaler bzw. extrazellulärer Coil-Regionen ist bei membranintegralen Proteinuntereinheiten an eine festgesetzte Reihenfolge gebunden. Nach einer solchen Region schließt sich in den meisten Fällen ein hydrophober helikaler Abschnitt an. Grundlegende Kriterien wie eine gute Fettlöslichkeit, Anzahl, Länge und Position der TM-Segmente sowie die Aminosäurezusammensetzung der Loop-Region bestimmen die Vorhersage in einem hohen Maß [Hil03].

4.7.2 Analyse

Die Vorhersage der Abfolge von Sekundärstrukturelemente ist von zentraler Bedeutung für das Erstellen einer 3D-Struktur. Zur Beurteilung angesprochener Strukturebenen von HP0565, sollte es möglich sein, Informationen mit Hilfe eines 2D-Modells und den Messdaten aus dem SMFS-Experiment auf eine 3D-Ebene zu verbessern. Damit eine exakte Aussage über die 2D-Struktur von HP0565 getroffen werden konnte, wurden zunächst geeignete bioinformatische Tools für eine Vorhersage der 2D-Topologie in Bezug auf Membranproteine ausgewählt: TMHMM 2.0, HMMTOP 2.0, MEMSAT3, ConPred II, SOSUI und TopPred. Die Analyse entspricht etwa einer Metavorhersage, die mehrere Vorhersageergebnisse kombiniert, um eine anzunehmende Fehlerquote einzelner Vorhersagen zu minimieren. Die Funktionsweise aller hier erwähnten Programme zielt auf eine Identifizierung von membraninternen- und externen Strukturen und soll an dieser Stelle etwas näher erläutert werden.

TMHMM 2.0 basiert auf einem Hidden-Markov-Modell, welches transmembrane Regionen auf Grundlage der Hydrophobizität vorhersagt. Darüber hinaus liefert das Programm Informationen über die Orientierung der einzelnen Domänen und somit auch über die des gesamten Proteins. Unterteilt werden die Abschnitte in „inside“, „outside“ und „transmembrane“ je nach zugewiesenem Hydrophobizitätswert für jede einzelne Aminosäure [SMR03]. Eine Vorhersage über **HMMTOP** basiert auf dem Grundsatz, dass die Topologie von TM-Proteinen durch eine maximale Abweichung der Aminosäurezusammensetzung der Sequenzsegmente bestimmt werden kann. HMMTOP erlaubt im Gegensatz zu TMHMM 2.0 eine Einteilung der fünf Gebiete, die als Grundlage für die strukturelle Beschreibung für HMM dienen (inside, outside, inside helix tail, outside helix tail, membrane helix) [TI01]. **MEMSAT3** wiederum erweist sich als effektive Methode für die 2D-Topologie Vorhersage. Der Algorithmus arbeitet auf Basis eines neuronalen Netzwerkes. Für eine wirksame Trennung von globulären -und transmembranen Proteinen wurde ein spezifisches neuronales Netzwerk trainiert, mit 399 Eingaben, 7 verdeckten Einheiten und lediglich einer Ausgabe. Zusätzlich wurde das Netzwerk mit einem Set, bestehend aus 134 transmembranen Proteinen und 416 zufällig ausgewählten globulären Proteinen trainiert. Für jedes einzelne der 184 Proteine konnte dann eine positionsspezifische Matrix berechnet werden [Jon07]. **ConPred II** ermöglicht eine Prognose von Anzahl und Lokalisation von membranständigen Bereichen. Unter ConPred II werden hauptsächlich zwei Systeme, zum einen ConPred-elite und zum anderen ConPred-all, zusammengefasst. Der gesamte Vorhersageprozess unterteilt sich in zwei aufeinanderfolgende Teilprozesse. Zu Beginn durchlaufen die Sequenzen das ConPred-elite Programm. Konnte beim ersten Durchlauf keine Topologie für die Sequenzen erstellt werden, durchlaufen diese anschließend ConPred-all für eine Vorhersage. Als Datensatz dient ein Set aus 138 prokaryotischen und 93 eukaryotischen Sequenzen mit bereits beschriebenen transmembranen Topologien [AMM04]. Die Anwendung **SOSUI** ist in der Lage anhand der vorliegenden Aminosäuresequenz zwischen Membran-und globulären Proteinen zu unterscheiden. Die Vorhersage von Membranproteinen schließt eine Bestimmung von transmembranen Helices mit ein. Das System stützt sich auf drei grundlegende Annahmen. Zunächst wird vorausgesetzt, dass in der Topologie vom Membranproteinen mindestens eine alpha-helikale Domäne vorkommt. Eine zweite Annahme beschreibt die Möglichkeit des Vorhandenseins von hydrophilen, wasserlöslichen transmembranen Bereichen in multispanning TM-Proteinen, vorausgesetzt deren Hydrophobizität entspricht auch den hydrophoben Segmenten von löslichen Proteinen. Die Software bezieht in die Vorhersage eine Reihe von physikochemischen Parametern ein, wie z.B den Hydrophobizitäts Index von Kyte und Doolittle, den Amphiphilizitäts Index von Aminosäuren und die Länge der Anfragesequenz [HBCM98].

Die Ausgabe der Ergebnisse der definierten Vorhersageprogramme wurde nach deren Möglichkeiten unterschiedlich visualisiert. Um eine hinreichende Interpretation der Resultate vorzunehmen, mussten die vorhergesagten Abschnitte, einheitlich mit der Einteilung nach: „transmembrane helix“, „loop (membrane-inside)“ und „loop (membrane-outside)“ in der Sequenz gekennzeichnet werden. Auf diese Weise konnte die Sequenz von HP0565 mit einer Länge von 215 residues, zeilenweise im Zusammenhang mit den Einzelergebnissen, auf gemeinsam vorhergesagte Abschnitte analysiert werden. Jedoch traf selten der Fall einer 100% Übereinstimmung der Vorhersagen ein. Diejenigen Bereiche, die beim Vergleich der einzelnen Ergebnisse keine genaue Sekundärstrukturzuordnung aufgrund von nicht eindeutig einzuordnenden Aminosäuren zuließen, wurden je nach Tendenz des Auftretens bestimmter Aminosäuren entweder zum Bereich Helix oder Loop bzw. coil (extern oder intern) gezählt. Die folgende Abbildung zeigt schematisch die resultierende Abfolge von Sekundärstrukturelementen für HP0565, die sich durch Überlagerung der Ergebnisse aus allen Anwendungen für die Vorhersage ergeben hat (Abbildung 4.7).

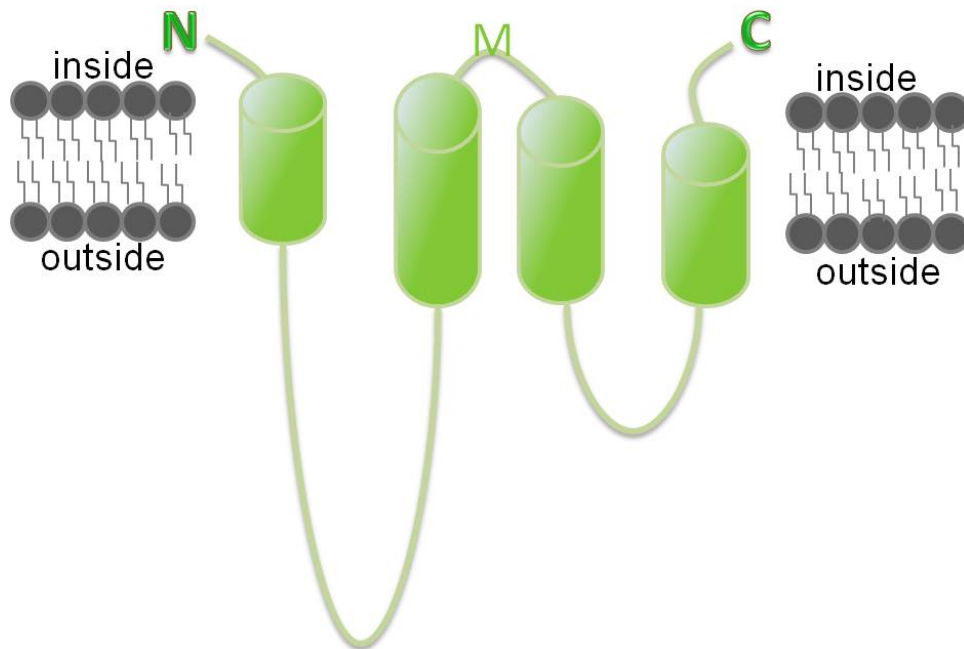


Abbildung 4.7: Topologie von HP0565. Resultierende Topologie von HP0565. Zu sehen sind vier transmembrane Bereiche und vorhergesagte inside/outside Regionen. Links und Rechts wird die hydrophobe Umgebung in Form der Lipiddoppelschicht angedeutet. Zweite und Dritte Helix werden nur über die Aminosäure Methionin miteinander verbunden.

Die Informationen, die sich nun über die Abfolge und die Länge spezifischer Sekundärstrukturelemente aus der 2D-Struktur des Proteins ableiten lassen, sollten hinsichtlich der Beurteilung einer Vorhersage der 3D-Struktur hilfreich sein. Da die Vorhersagewahrscheinlichkeiten zwischen 40% und höchstens 70% (Vorhersage über neuronale Netze) der verwendeten Programme schwanken und die Aminosäuren entsprechend ihrer Tendenz, einer bestimmten Struktureinheit zugehörig zu sein, beurteilt werden, sollte die dargestellte 2D-Struktur von HP0565 als möglich angenommen werden, jedoch nicht als endgültig. In Bezug auf die 3D-Ebene und der Auswertung der Messdaten, wird jedoch im weiteren Verlauf der vorliegenden Arbeit deutlich, dass in der Topologie mehr als nur ein primitives Muster für die 3D-Struktur steckt.

In der folgenden Übersicht (Tabelle 4.4) sind die benannten Bereiche mit ihrer Länge, entsprechend der Sequenz, aufgeführt.

Bereich	Länge [aa]
inside	1-18
transmembrane helix	19-33
outside loop	34-101
transmembrane helix	102-126
inside (Methionin)	127
transmembrane helix	128-150
outside loop	151-166
transmembrane helix	167-185
inside	186-215

Tabelle 4.4: Topologische Bereiche

4.8 Vorhersage der 3D-Struktur von HP0565

4.8.1 Grundlagen

Die bioinformatische Beschreibung des hypothetischen Proteins beabsichtigte vorrangig eine Vorhersage eines 3D-Modells von HP0565, welches im günstigsten Fall ermöglicht, spezifische Entfaltungswege von HP0565 besser zu verstehen. Die auf Sequenzhomologie beruhende Proteinmodellierung ist eine sehr hilfreiche und vor allem schnelle Methode zur Strukturvorhersage von Proteinen. Dabei wird die Erkenntnis, dass verwandte Proteine einer Proteinfamilie, die eine hohe Ähnlichkeit ihrer Aminosäuresequenzen aufweisen, meist auch sehr ähnliche Proteinfaltungen haben. Während einer solchen Modellierung wird zunächst die Aminosäuresequenz des zu modellierenden Proteins mit der Sequenz des Referenzproteins verglichen, indem ein paarweises oder bei mehreren Proteinen ein multiples Sequenzalignment durchgeführt wird. Liegen die Sequenzidentitäten über 70%, so können die modellierten Strukturen sehr genau vorhergesagt werden. Hingegen können Sequenzidentitäten unter 30% zu Problemen bei der Modellierung führen. Die Sequenzidentitäten von strukturell konservierten Regionen liegen jedoch häufig über denen der weitaus weniger konservierten Schleifen (Loops). Somit können also die Schleifen das Maß der Identität der gesamten Sequenz merklich beeinflussen. Um strukturell konservierte Regionen in den Referenzproteinen zu identifizieren, wird ein Strukturalignment der Aminosäuresequenzen auf Basis der Sekundärstruktur durchgeführt. Anschließend wird die zu modellierende Sequenz dem Template angepasst und die räumlichen Koordinaten der strukturell konservierten Regionen übertragen. Die Koordinaten der Loops werden meist von ähnlichen Bereichen anderer Proteinstrukturen übernommen. Die räumliche Ausrichtung der Seitenketten der einzelnen Aminosäuren wird in konservierten Bereichen wie in den Templates beibehalten und für alle nicht konservierten Seitenketten wird die statistisch wahrscheinlichste Position gewählt. Berechnungen, die zur Energie-minimierung des Modells führen sowie die Prüfung der strukturellen Relevanz des Proteinmodells schließen den Vorgang der Homologiemodellierung ab [Hil03].

4.8.2 Analyse

Für eine homologiebasierte Modellierung der 3D-Struktur von HP0565 musste zunächst eine geeignete Auswahl der Templates bzw. Referenzproteine getroffen werden. Hierbei wurden die Musterstrukturen nicht zufällig ausgewählt, sondern nach signifikanten Ähnlichkeitsmaßen aus der Ähnlichkeitssuche. In Anbetracht der Ergebnisse aus der BLAST-Suche waren die Sequenzen, die die höchste Identität zu HP0565 aufwiesen, ebenfalls als hypothetisch hinterlegt und ohne strukturelle Informationen. Eine erste Auswahl fiel auf die Proteine YKGB-Ecoli und YKGB-PASHA der Spezies *Escherichia coli* *Pasteurella haemolytica*. Diese beiden Organismengruppen lieferten anfangs nützliche Sequenz- und Strukturhinweise, die sich gegebenenfalls auf HP0565 übertragen lassen sollten. Für die Modellierung der Strukturen wurden primär die Datenbanken ModBase und der Protein Structure Prediction Server (PS)² genutzt. Das Angebot von bioinformatischen Tools, die eine 3D-Struktur auf der Grundlage von Sequenzähnlichkeiten vorhersagen ist groß. ModBase ist eine Datenbank, welche annotierte Proteinstrukturmodelle verwaltet. Die Modelle werden über ModPipe abgeleitet. Das Tool ModPipe stützt sich für eine Modellierung auf die Programme PSI-BLAST und MODELLER. ModBase beinhaltet nicht nur bereits vorhergesagte Proteinstrukturen, sondern auch Alignments und Faltungsphänomene basierend auf der Modellierung der Modelle. Für eine Vorhersage bezieht das Programm ausschließlich Informationen aus theoretisch kalkulierten Modellen ein und experimentelle Fakten werden ignoriert [PEB04]. Die Arbeitsweise des Protein Prediction Servers basiert ebenfalls auf der Methode der Homologiemodellierung. Die Effizienz dieser Anwendung ist auf eine Kombination verschiedener Lösungsstrategien zurückzuführen, welche zum einen die Auswahl des Templates (über PSI-BLAST und IMPALA) und zum anderen die Ausführung des Alignments mit dem Ziel-Template (über PSI-BLAST, IMPALA und T-Coffee) umfassen. Während der Ausführung von (PS)² werden vier Schritte nacheinander durchgeführt. Im ersten Schritt erfolgt die Auswahl des Templates, folgend das Alignment bzw. die Anordnung des Ziel-Templates, das Erstellen eines Modells und schließlich die Bewertung und Visualisierung des Modells. Diese Schritte werden so oft wiederholt, bis eine zufriedenstellende Struktur erreicht wird. Für das Modellieren eines Modells greift (PS)² auf das Tool MODELLER zurück [CCJKJM06]. Das Programm ermöglicht dem Anwender selbst ein Template durch die Angabe der PDB Identifikationsnummer zu wählen bzw. über eine Suche, basierend auf PSI-BLAST und IMPALA, ein Template für die Modellierung zu identifizieren.

Eine direkte Abfrage eines Modells für HP0565 über Modbase erwies sich als ergebnislos. Dagegen konnte ein Modell von YKGB-Ecoli K12 und YKGB-PASHA erstellt werden. Dabei wurde automatisch das Template mit der PDB-Identifikationsnummer 2iw3A aus-

gewählt. Hinter dieser ID verbirgt sich der Elongationsfaktor eEF3, eine ATPase, die im Zusammenspiel mit den Faktoren eEF1A und eEF2 eine entscheidende Rolle im Übersetzungszyklus von Pilzen spielt. Dieses Enzym ist in der Lage ATP in ADP und Phosphat aufzuspalten. Die in der PDB hinterlegte Struktur mit der Identifikationsnummer 2iw3A stellt die Struktur des Faktors eEF3 (chain A) aus *Saccharomyces cerevisiae* dar. Mit einer Sequenzlänge von 973 residues war es nicht unwahrscheinlich, dass durch die Größe des Faktors eine Struktur mit den Anfragesequenzen von YKGB-Ecoli K12 und YKGB-PASHA zustande kam. Die modellierte Struktur von YKGB-Ecoli K12 bleibt jedoch fraglich bzw. wurde hierbei die Genauigkeit der Struktur in Frage gestellt, da lediglich eine Identität der Sequenzen von 15% vorherrschte. Das Modell von YKGB-PASHA wurde zunächst nicht weiter beurteilt, da ein zu kurzer Abschnitt aligniert wurde und diese Struktur nicht aussagekräftig genug war. Daraus wurde an diesem Punkt erkannt, dass sich ausgehend von diesen beiden Modellen, keinerlei strukturelle Präferenzen auf die mögliche Struktur von HP0565 übertragen ließen. Die Abfrage über $(PS)^2$ lieferte ein Modell von HP0565. Auch hier diente der Faktor eEF3 als Template. Jedoch konnten nur sehr kurze Abschnitte von der Sequenz von HP0565 auf die des Templates übertragen werden. Aus diesem Grund ist dieses Modell nicht signifikant genug, um es an dieser Stelle vorstellen zu können. Infolgedessen mussten nun neue mögliche Musterstrukturen gefunden werden, die ein Modellieren eines Modells möglich machen könnten.

Die „Kyoto Encyclopedia of Genes and Genomes“ (KEGG) [KG00] ist eine sehr umfangreiche, öffentlich zugängliche Datenbanksammlung zu Pathway relevanten Informationen, die zu den Angeboten des japanischen GenomeNet Datenbank Services (GenomeNet) zählt. Diese Datenbank umfasst das Wissen in Bezug auf die molekularen Interaktions-Netzwerke biologischer Prozesse (PATHWAY Datenbank), das Wissen über den Bereich der Gene und Proteine (GENES/SSDB/KO Datenbanken) und die Welt der chemischen Verbindungen und Reaktionen (COMPOUND/REACTION Datenbanken). Ziel der Annotationen ist es, die beschriebenen Gensequenzen in Verbindung mit orthologen Genbezeichnern (ortholog identifiers) zu bringen. Diese stehen für einander entsprechende Gene gleicher Funktion, die (auch aus den verschiedenen Organismen stammend) als Gene evolutionär verwandten Ursprungs zu betrachten sind [KG00]. Eine Suche in der beschriebenen KEGG-Datenbank zielte auf ein Identifizieren von paralogenen, bereits beschriebenen Proteinen, der Spezies *Helicobacter pylori* (Stamm 26695). Da der Versuch eine Struktur von HP0565 mit den bisher verwendeten Templates zu modellieren, zu keiner aussagekräftigen Struktur führte, wurden in nachfolgenden Schritten Musterstrukturen der strukturell beschriebenen Proteine des Stammes für HP0565 eingesetzt. Betrachtet wurden dabei translatierte Bereiche des Genomabschnittes 570001 bis 600000

Basenpaare, wobei der für HP0565 kodierende Bereich in der Region von 597293 bis 597940 lokalisiert ist.

Das gesamte Genom des Stammes 26695 besitzt eine circuläre Gestalt und ist 1,6 mBasenpaare groß. Für ein weiterführendes modellieren wurden folgende Proteine des Stammes 26695 betrachtet (Tabelle 4.5).

Protein	Orthologie	Template PDB-ID
HP0561	3-oxoacyl-reductase [EC:1.1.1.100]	Struktur auf Template 1dohA modelliert
HP0558	3-oxoacyl-synthase II [EC:2.3.1.179]	Struktur auf Template 1j3nA modelliert
HP0544	cag pathogenicity island protein 23	Struktur auf Template 1e9rA modelliert
HP0566	diaminopimelate epimerase [EC:5.1.1.7]	Struktur auf Template 2otnA modelliert
HP0663	chorismate synthase [EC:4.2.3.5]	PDB-Id 1UM0A

Tabelle 4.5: Annotierte Proteine des Stammes 26695 [KG00]. Die dargestellte Tabelle zeigt einen Datensatz der bereits beschriebenen Proteine des Stammes 26695 und deren Templatestrukturen, auf die diese modelliert werden konnten. Die Angabe der PDB-Identifikationsnummer bei HP0663 verweist darauf, dass dieses bereits mit einer 3D-Struktur in der Proteindatenbank abgelegt ist.

Es schloss sich eine Erstellung der Strukturen auf Basis der Homologiemodellierung mit dem Tool „Protein Prediction Server (PS)2“ an. Jedoch wurden nur zwei der Templates erkannt (1dohA und 1UM0A), die verwendet werden konnten. Auch hier war die Qualität aufgrund von geringen Struktur- und Sequenzidentitäten zwischen Ziel- und Musterstruktur nicht zufriedenstellend. Auch für das bereits charakterisierte Membranprotein HP0567, welches in der Nähe des translatierten Bereiches von HP0565 liegt, wurde über die Proteinstrukturdatenbank „ModWeb“ ein 3D-Modell angefordert. Allerdings war der Sequenzbereich, der mit dem der Musterstruktur align wurde zu kurz, um bedeutsame Informationen abzuleiten. Da funktionelle Untereinheiten (Domänen) und konservierte Bereiche oftmals mit strukturellen Konservierungen einhergehen, wurden nochmals Datenbanken auf eventuell vorhandene Motive von HP0565 untersucht. Allerdings erwies sich die Suche als wenig hilfreich, denn außer der DUF417 Domäne, welche uncharakterisierte Proteine beschreibt, ließen sich keine anderen Domänen, Motive oder aktive Bindungsmöglichkeiten des Proteins finden. Schlussfolgernd und unter Einbeziehung sämtlicher Eindrücke aus der Strukturanalyse wird deutlich, dass mit den zur Verfügung stehenden Prinzipien aus der bioinformatischen Beschreibung und den experimentellen Methoden es momentan nicht möglich ist ein aussagekräftiges Modell von HP0565 vorzustellen. Die Ergebnisse aus den Vorhersagen lassen die Modelle in einem hohen Maß anzweifeln. Für die folgende Kurvenanalyse bedeutet dies, dass strukturell eine Rückkehr zur 2D-Struktur

von HP0565 erfolgen muss, um nützliche Informationen über die Entfaltung von Sekundärstrukturelementen zu erhalten.

Kapitel 5

Datenanalyse

5.1 Betrachtungen zum Experiment und experimentelle Eckdaten

Durch ein am hypothetischen Protein durchgeführtes SMFS-Experiment sollte es möglich werden Einblicke in strukturelle Veränderungen zu erlangen, ohne das nützliche Hinweisse aus der dreidimensionalen Struktur des Proteins derzeit zur Verfügung stehen. Lediglich die Topologie des Proteins hat hier eine richtungsweisende Funktion im Hinblick auf die Identifizierung von Stabilitätsbarrieren. Durch eine Aufzeichnung und anschließende Auswertung von Kraft-Abstandskurven wird es möglich intra- und intermolekulare Wechselwirkungen eines Proteins als individuellen Fingerprint darzustellen. Über diesen Fingerprint lassen sich verschiedene Informationen ableiten, die eine dreidimensionale Anordnung aller Atome im Raum nicht in jedem Falle erfordert. Wie auch bei jedem Experiment, spielt die Probenpräparation eine entscheidende Rolle, um ein „erfolgreiches“ Experiment zu gewährleisten. Die Präparation ist abhängig von der Art des Biomoleküls und dessen physiologischen Ansprüchen an die vorherrschende Umgebung. Nicht selten werden für die Auswertung und Gewährleistung der Signifikanz der Kraft-Abstandskurven mehrere hundert bzw. tausend solcher Kurven aufgenommen.

Bei Membranproteinen ist die Probenvorbereitung für ein SMFS-Experiment im Gegensatz zu globulären Proteinen ein komplexer, aufwendiger Vorgang, der mit mehreren Schritten verbunden ist. Dabei sind spezielle zellbiologische Verfahren erforderlich. Wer Membranproteine reinigen oder untersuchen will, bringt sie besser zuerst in Lösung. Integrale Membranproteine lösen sich mit Seifen. Periphere Membranproteine d.h. mit Membranen assoziierte Proteine, die keine Transmembranhelices besitzen, lösen sich oft schon in Puffern hoher oder niedriger Ionenstärke oder hohem pH-Wert. Durch das Verfahren der Solubilisierung sollen Konformation und Funktion des Membranproteins so gut wie

möglich erhalten bleiben. Bei integralen Membranproteinen ist dies oft ein Problem. Des-
sen Lösung kann aber mit Glück nicht nur den Weg zur Reinigung öffnen, sondern auch
Einblicke in die Funktion gewähren. Beim Solubilisieren lagern sich die Seifenmoleküle
mit ihren hydrophoben Resten an die hydrophoben Stellen (vor allem Transmembranbe-
reiche) des Proteins an und verdrängen dabei teilweise die Phospholipide [HL09] (Abbil-
dung 5.1). Lagern sich genügend Seifenmoleküle an, geht das Membranprotein in Lösung.
Solubilisierte Membranproteine sind also Komplexe aus Seife, Phospholipid und Protein,
wobei die Anteile der Komponenten von der Zusammensetzung des verwendeten Puffers
abhängen. Der Anteil der Seifen- und Phospholipidmoleküle am Komplex liegt in der Re-
gel zwischen 10-50%. Jedes Membranprotein benötigt eigene Bedingungen, um in nativer
Form in Lösung zu gehen. So spielen der pH-Wert, der Puffer und die Ionen eine wichtige
Rolle. Mit Hilfe des Solubilisierens, wird es möglich gezielt Mizellen hervorzubringen
[HL09]. Mizellen sind thermodynamisch stabile Aggregate grenzflächenaktiver Moleküle.
Die hydrophoben Teile der Seifenmoleküle befinden sich im Inneren der Mizellen, die
hydrophilen Reste wechselwirken mit dem wässrigen Medium.

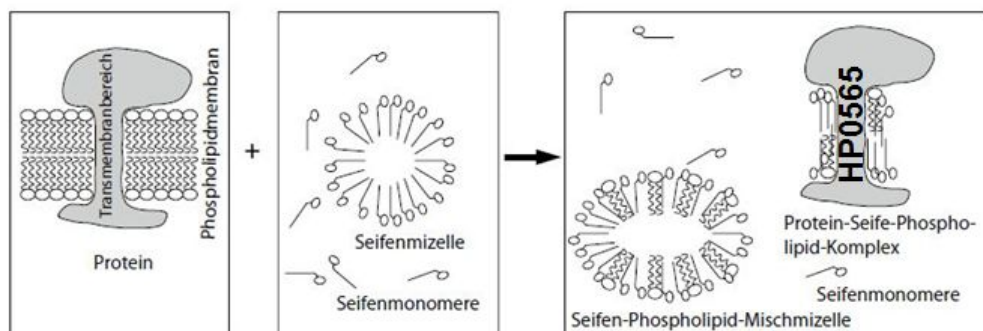


Abbildung 5.1: Solubilisierung von integralen Membranproteinen (modifiziert nach [HL09])

Beim SMFS-Experiment von HP0565 war es möglich solche Aggregate mit dem Prote-
in hervorzubringen. Diese Charakteristik bei der Probenvorbereitung, machte es mitunter
möglich das hypothetisch beschriebene Protein als Membranprotein vorzustellen. Globu-
läre Proteine wiederum weisen nach außen hin hauptsächlich hydrophile Domänen auf,
sodass keine hydrophoben Bereiche bedeckt werden müssen und die Anordnung in Mi-
zellen nicht möglich ist. Die Probenexpression erfolgte am FMP Berlin. Das eigentliche
Experiment wurde jedoch am BIOTEC Dresden durchgeführt.

Nachdem die Umgebung mit dem Protein HP0565 vorbereitet wurde, konnte das eigentliche Entfaltungsexperiment (Abbildung 5.2), gemäß dem erwähnten Prinzip einer erzwungenen mechanischen Entfaltung des Biomoleküls, durchgeführt werden. Durch das Heranfahren der Tastspitze an die Oberfläche, wurden nacheinander terminale Enden des Proteins berührt und bei Anwendung einer geringen Kraft (ca. 300 pN), Proteinketten aus der Umgebung entfernt. Der noch vorhandene Teil des Aggregates bleibt bestehen. Bei einer ausreichenden Zugkraft der Feder (ca. 400 pN) entfalten die einzelnen strukturellen Elemente von HP0565. Das resultierende Kraftspektrum ermöglicht eine Zuordnung einer mechanischen Stabilität. Nach der Detektion des Kraftspektrums wird die gleiche Stelle der Oberfläche mit den sogenannten „patches“ abgebildet, um sicher zu stellen, dass das Spektrum nur durch ein Fragment beschrieben wird. Die Interpretation einer Kraftkurve lässt sich wie folgt darlegen: Wenn sich die AFM Spitze von der Oberfläche entfernt, werden zunächst die unspezifischen Kräfte zwischen beiden Oberflächen überwunden. Dieses Verhalten lässt sich durch die Signalgebung in (a) erkennen. Die Adhäsionsmaxima liegen meist innerhalb eines Abstandes von 0-15 nm. Je nachdem welches terminale Ende an die Tastspitze adsorbiert (in dem Falle d. N-terminale Ende), wird dieses weiter gedehnt und es bildet sich eine steigende nichtlineare Kraft aus. Wobei zu Beginn der Betrachtung der FD-Kurven nicht klar war, welches terminale Ende an die Tastspitze adsorbiert. Ab einer bestimmten Kraft entfaltet sich, entsprechend des Beispielkraftmaximums von HP0565, z.B die erste Helix (b). Nach diesem ersten Entfaltungsvorgang fällt die Kraft wieder auf null ab. Die Polypeptidkette wird nun weiter gestreckt und die Kraft steigt weiter an, sobald das Peptid gedehnt wird. Ab einer bestimmten Kraft entfalten sich nachfolgende Sekundärstrukturelemente (c-f) (Abbildung 5.2 C).

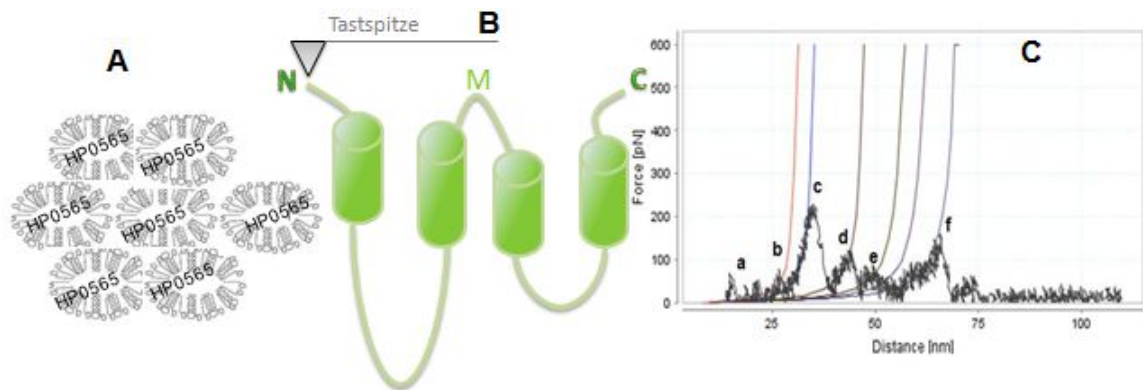


Abbildung 5.2: Kontrolliertes Entfalten von HP0565 (modifiziert nach [nan10] und [HL09]). **A**, Abbildung von Mizellenaggregaten, die aus einem Komplex von Seife-Phospholipid und dem Protein HP0565 vorliegen. **B**, Auswahl eines Protein-Komplexes. Die Tastspitze nähert sich der Oberfläche an und adsorbiert im Beispiel an das N-terminale Ende. Eine mechanische Entfaltung wird hier und in folgenden Ausführungen durch die Topologie dargestellt und geschildert. **C**, dargestellt ist eine aufgenommene Kraft-Abstandskurve von HP0565 mit WLC-Fit. Diese FD-Kurve ist nach dem Alignierungsvorgang etwa 100nm lang und zeigt detaillierte Einsichten in die Entfaltung des Proteins.

Dieser graduelle, nichtlineare Anstieg der Kraft wird im Falle von HP0565 mit dem WLC-Modell gefittet. Die Konturlänge des gedehnten Polypeptids ist der einzig freie Parameter, der in diesem Fit verändert werden kann. Die Polypeptidlänge, die aus diesem Fit hervor geht, kennzeichnet die ungefähre Lage von möglichen Stabilitätsbarrieren innerhalb des Proteins. Entsprechend diesem Fit, wurde der bearbeitete Datensatz auf mögliche Entfaltungswege hin untersucht. Der komplette Datensatz an Kraft-Abstands Kurven umfasste mehrere hundert Kurven, die innerhalb eines bestimmten Zeitintervalls aufgenommen wurden, wobei der Anteil von nicht repräsentativen Kurven selektiert werden musste. Dies realisierte vorwiegend die bereitgestellte Software [nan10] nach bestimmten Ähnlichkeitsmaßen. Analog zu anderen Methoden nimmt die Analyse und Auswertung der Daten einen Großteil der wissenschaftlichen Arbeit ein. Durch automatisierte Auswerteverfahren kann der beträchtliche zeitliche Aufwand signifikant minimiert werden. Dabei steht neben der Zeitersparnis auch die Objektivität und Vergleichbarkeit im Vordergrund. Der bereinigte Datensatz wurde anschließend nach bestimmten Auswahlkriterien geprüft. Artefakte, wie unspezifische Kraftmaxima, bestimmten die Auswahl von nicht aussagekräftigen FD-Kurven. Weiterhin war die Menge an Daten durch eine merklich hohe Variabilität in den Kurvenmustern gekennzeichnet. Nach intensiver Betrachtung des Datensatzes blieben etwa 800 Kurven für anstehende Analysen übrig. Im Laufe der Untersuchungen minimierte sich dieser weiter.

5.2 Anwendung verschiedener Strategien für die Auswertung der FD-Kurven von HP0565-Datenfilterung-

Die angesprochene Variabilität im Datensatz von mehreren hundert aufgenommenen FD-Kurven erforderte die Anwendung verschiedener Lösungsstrategien, im Hinblick auf eine Stabilitätsanalyse. Denn beim bloßen Betrachten aller Kraftmaxima konnten keine sinnvollen Rückschlüsse auf membranstabilisierende Bereiche gezogen werden. Die Kurven liefern lediglich nur Aussagen über die gemessene Kraft, wenn sich die Tastspitze von der Oberfläche der Membran entfernt und sich nachfolgend bei steigender Kraft membraninterne Strukturen entfalten. Zusätzlich werden Abstände zwischen entfalteteten, aufeinanderfolgenden Fragmenten visualisiert. Darum mussten für die Bewertung des Datensatzes, verschiedene Strategien verfolgt werden. Anfänglich mussten mögliche Entfaltungsmuster anhand der festgelegten 2D-Topologie für N- und C-Terminus gefunden werden. Hierfür war eine Betrachtung beider Termini erforderlich, da zu Beginn der Analyse unklar war welches terminale Ende an die Tastspitze adsorbiert. Mit diesem Hintergrund erschloss sich die Fragestellung: Wie würde eine Kraftabstandskurve aussehen, wenn das Protein ausgehend vom C-terminalen Ende oder vom N-terminalen Ende entfaltet? In diesem Zusammenhang sollte geklärt werden, welche Sekundärstrukturelemente sich in welcher Reihenfolge entfalten oder ob diese immer einzeln oder auch paarweise zusammen in einem Entfaltungsschritt gelöst werden. Als Antwort hierauf entstanden unterschiedliche Gruppen von Sekundärstrukturentfaltungsmustern, die im Folgenden vorgestellt sind. Ebenso sollte es möglich sein anhand dieser Entfaltungsmuster eine Signalzuordnung vorzunehmen, die die Absicht verfolgte, main- und sidepeaks im Kurvensatz zu detektieren. Main-peaks kennzeichnen Signale, die innerhalb des Datensatzes gehäuft bzw. in jedem Entfaltungsweg zu erkennen sind. Side-peaks tauchen hingegen nur vereinzelt auf. Durch die Detektion genannter main-Peaks können Hauptentfaltungsschritte des Proteins beschrieben werden. Hingegen kennzeichnen side-Peaks Nebenentfaltungsschritte. Durch sequenziell bestimmte Entfaltungsbarrieren konnte der Datensatz von etwa 300 Kurven interpretiert werden. Für einen Entfaltungsvorgang spezifischer Sekundärstrukturelemente, wurde lediglich der Fall einer kompletten Entfaltung von Helix-Domänen beurteilt. Über Abstände in den Konturlängen der jeweiligen Kraftkurve erfolgte eine Einordnung in einen möglichen Entfaltungsfall. Diese Klassifizierung verfolgte die Absicht, Aussagen über den terminalen Weg durch den das Protein hauptsächlich aus der Membran mechanisch entfernt wurde, treffen zu können. Bei der vorliegenden Klassifizierung wurde darauf geachtet, dass Peaks, die einer unspezifischen Wechselwirkung zwischen Cantile-

ver und Oberfläche in einem Bereich von ca. 0-15 [nm] zuzuordnen waren, nicht beachtet wurden. Abstände zwischen erstem spezifischen Kraftmaximum und dem nachfolgenden wurden ebenfalls zunächst vage beurteilt. Denn die Konturlängen variieren in Abhängigkeit von der Länge der terminalen inside-Bereiche (N-terminal: 1 bis 18 [aa], C-terminal: 1 bis 29 [aa]), je nachdem wo die Tastspitze ansetzt.

Die folgende Tabelle (Tabelle 5.1) soll einen Überblick über aufgestellte Entfaltungsfälle geben.

Fall	Terminus	Konturlänge [aa]	Reihenfolge der Entfaltung
1	N	102, 128, 167, (185)	Helice entfalten einzeln, nacheinander
2	N	102, 167, (185)	1 H. einzeln, 2-3 Helix zusammen, 4 H. einzeln
3	N	128, 167, (185)	1-2 H. zusammen, 3 H. und 4 H. einzeln
4	N	102, 128, (185)	1 H. und 2 H. einzeln, 3-4 H. zusammen
5	N	167, (185)	1-3 H. zusammen, 4 H. einzeln
6	N	102,(185)	1 H. einzeln, 2-4 H. zusammen
7	N	unspezifisches Signal	Komplette entfalten aller Elemente in einem Schritt
8	N	128, (185)	1-2 H. zusammen, 3-4 H. zusammen
9	C	65, 89, 182, (197)	Helice entfalten einzeln, nacheinander
10	C	89, 182, (197)	1-2 H. zusammen, 3 H. einzeln, 4 H. einzeln
11	C	65, 89, (197)	1 H. einzeln, 2 H. einzeln, 3-4 H. zusammen
12	C	182, (192)	1-3 H. zusammen, 4 H. einzeln
13	C	65, 182, (192)	1 H. einzeln, 2-3 H. zusammen, 4 H. einzeln
14	C	unspezifisches Signal	Komplettes Entfalten aller Elemente in einem Schritt
15	C	65, (197)	1 H. einzeln, 2-4 H. zusammen
16	C	89, (197)	1-2 H. zusammen, 3-4 H. zusammen

Tabelle 5.1: Mögliche Entfaltungsfälle. Die Tabelle zeigt die auf der Sequenz von HP0565 beruhenden Endpunkte der Entfaltung von vorhandenen Sekundärstrukturelementen. Die erwähnten Endpunkte sind mit den Entfaltungslängen definiert. Die in den Klammern aufgeführten Längen kennzeichnen, ausgehend von Terminus, Signale, welche durch die Enden des letzten membranständigen Elementes entstehen.

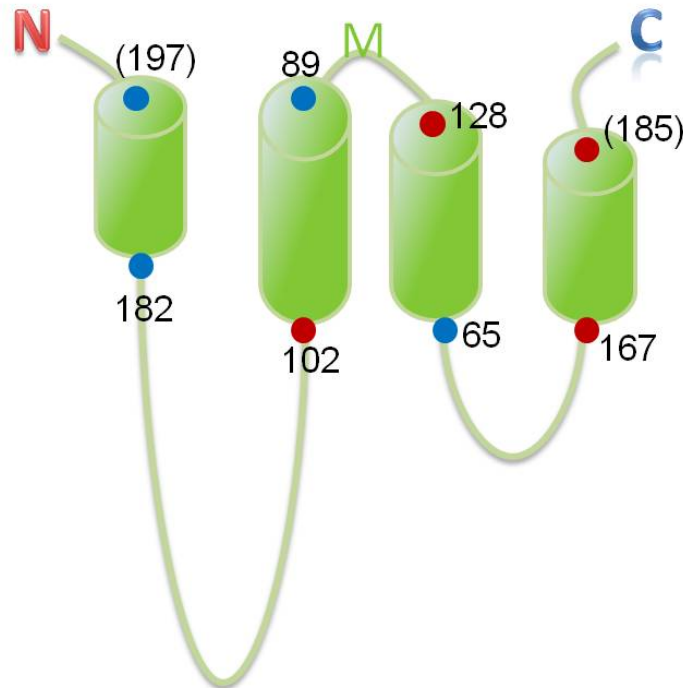


Abbildung 5.3: Topologie mit gekennzeichneten Endpunkten möglicher Entfaltungen. Die dargestellte Topologie von HP0565 vermittelt die terminalen Endpunkte möglicher Entfaltungsschritte. Blaue Punkte kennzeichnen die C-terminalen Endpunkte und rote Punkte mögliche N-terminale Entfaltungen. Die in Klammern stehenden sequenziellen Entfaltungspunkte beschreiben die letzten, möglichen terminalen Entfaltungspunkte.

Es gelang das Set an 300 signifikanten FD-Kurven des Membranproteins HP0565 den aufgestellten Entfaltungsfällen zuzuweisen (Tabelle 5.1). Dabei war es möglich jedem charakteristischen Fall mindestens zwei Kurvenverläufe zuzuordnen. Jedoch konnten ebenso viele Kurven einem charakteristischen N-terminalen Entfaltungsfall wie C-terminal zugeteilt werden. Aus diesem Grund war es an dieser Stelle nicht möglich genaue Aussagen darüber zu treffen, ob das Protein mit einer symmetrischen Lage der Sekundärstrukturelemente in den Mizellenaggregaten überwiegend C-terminal oder N-terminal aus dieser Umgebung entfernt wurde. Die Arbeitsweise der verwendeten Software [nan10] beeinflusste diesen Lösungsweg zusätzlich, denn das Programm sucht im zu analysierenden Datensatz nach gemeinsam auftretenden Kraftmaxima, welche durch das WLC-Modell gefittet werden. Beim Erstellen der experimentellen Daten bindet die Tasfeder zufällig an das Protein d.h. die Tastspitze der Tasfeder kann an jedem, sich nicht in der Membran befindlichen Element andocken. Somit ist nicht eindeutig festgelegt wo sich der eigentliche Anfang des Proteins befindet. Genau diese Thematik spiegelt sich auch in der Messreihe wider. Der Nullpunkt kann sowohl in den negativen als auch in den positiven Bereich verschoben worden sein. Eine Negative Verschiebung kann einer Fragmentanhaftung an der Tastspitze zugrunde liegen, wodurch eine künstliche Proteinsegmentverlängerung er-

zwungen wird [Cle06]. Somit treten relevante Signale erst später im Spektrum auf. Um eine exakte Analyse der Datensätze zu ermöglichen, muss ein Verfahren angewendet werden, mit dem eine exakte Kontaktbereichsbestimmung vorgenommen werden kann. Das unterschiedliche Maß an Verschiebung einzelner FD-Kurven beeinflusste den ersten Lösungsweg deutlich. Somit würden sich falsch-positive Ergebnisse hinsichtlich möglicher Entfaltungswege des Proteins ergeben, bei einer nach eigenen Gesichtspunkten festgelegten Analyse. Die erwähnte Positionsänderung der Kurven wurde jedoch bei der vorgenommenen Betrachtung zunächst nicht mit einbezogen.

In Folge dessen musste eine neue Lösungsstrategie aufgestellt werden, damit eine Stabilitätsanalyse möglich wird. Dabei sollten die Parameter des vorangegangenen Lösungsweges und zusätzlich die Verschiebung der Kraft-Abstandskurven mit eingeschlossen werden. Aus dem zuvor verwendeten Datensatz wurden für den folgenden Lösungsansatz nur die längsten FD-Kurven ausgewählt. Die Software bietet dem Anwender die Möglichkeit, die Daten nach dem statistischen Vorkommen gemeinsamer Peaks zu untersuchen. Dadurch konnten für weitere Schritte 50 Kurven, welche ein im arithmetischen Mittel auftauchendes Signal von 193 [aa] aufwiesen, ausgewählt werden. Diese FD-Kurven wurden ebenfalls wiederholt auf deren Repräsentativität geprüft, um sicher zu gehen, dass der beschriebene Lösungsansatz mit den besten Kurven realisiert werden konnte. Somit blieben nach einem Ausschließen nicht repräsentativer Kurven 35 verwendbare übrig. Das Set wurde anschließend durch die Informationen über eine Entfaltung entsprechend sequenzieller Punkte manifestierter helikaler Bereiche aus der 2D-Topologie (Tabelle 5.2) begutachtet. Dabei wurde eine terminale Betrachtung vorgenommen.

Signale N-terminal [aa]	(19)	102	128	167	(185)
Signale C-terminal [aa]	(30)	65	89	182	(197)

Tabelle 5.2: Signale entsprechend der 2D-Topology von HP0565. Die Tabelle beinhaltet sequenziell bestimmte Entfaltungspunkte manifestierter helikaler Bereiche. Die in Klammern aufgeführten Entfaltungspunkte, der N-terminalen und C-terminalen Betrachtung, wurden zunächst nicht mit in die folgende Beurteilung einbezogen.

Die terminalen Wechselwirkungen, welche eine Signalgebung im N-terminalen Bereich bei ungefähr 19 [aa] (distance) und C-terminal bei etwa 30 [aa] (distance) im Kurvenspektrum zeigen, blieben vorerst aus der erwähnten Problematik heraus unberücksichtigt. Jede Kurve wurde auf den Originalzustand vor dem Alignierungsverfahren über die Software [nan10] untersucht. Hierbei wurde der Bereich, um den der Kurvenverlauf positionsmäßig verschoben werden musste, abgezogen. Dieses Vorhaben zielte auf ein Umgehen möglicher Fehlzuordnungen von Kraftmaxima.

Definierte Konturlängen konnten somit anhand der im Voraus aufgestellten Entfaltungspunkte für jede Kurve verglichen und zugeordnet werden. Selten konnte eine 100%-ige Übereinstimmung zwischen neu berechneten Entfaltungslängen und den terminal manifestierten Entfaltungspunkten gefunden werden. Ein definierter Referenzbereich von plus/minus 10 [aa] zwischen neu definierter Konturlänge und der Länge einer Entfaltung aus der Topologie sollte dieser Problematik Abhilfe schaffen. Die Kurven wurden anschließend nach deren Ähnlichkeit zueinander betrachtet und in Abhängigkeit ihrer Verschiebung unterschiedlichen Gruppen zugeteilt. Bestimmend war hierbei die Lage der FD-Kurven nach dem alignieren. Die Zuordnung in die jeweiligen Gruppen bezweckte in Hinsicht auf eine Stabilitätsanalyse das Auffinden von möglichen Entfaltungswegen. Ein Entfaltungsweg kennzeichnet die Abfolge von Entfaltungsschritten, die in mehreren Kurven sichtbar werden. An diesem Punkt konnten weitaus mehr FD-Kurven durch eine Konturlängenzuordnung dem N-terminalen Entfaltungsfall zugeordnet werden. Als Hilfsmittel für eine exaktere Zuordnung diente eine weitere Software (DURIN). DURIN [Cle06] ist eine Weiterentwicklung des Programmes MPTV [Mül05] und wurde zur Bearbeitung und Interpretation von experimentell ermittelten SMFS-Datensätzen entwickelt. Dadurch sollte eine Aufarbeitung von Rohdatensätzen in Verbindung mit Proteindaten möglich werden, sodass ein Informationsverlust vermieden wird und verfälschte Informationen umgangen werden. Ein auf Matrizen basierender Alignmentalgorithmus schaffte eine Möglichkeit, mit der die Datensätze schneller und effektiver analysiert werden konnten. Im Zusammenhang mit dem zweiten vorgestellten Lösungsansatz sollten mit Hilfe der automatisierten Möglichkeiten des Programmes ähnliche Kurvenmuster in den Gruppen gefunden werden, um nachfolgend eine statistische Analyse von main- und side-Peaks zu gewährleisten. Resultierend wurden demnach eigene per Hand vorgenommen Analysen mit den Möglichkeiten einer umfassenden Datenanalyse der Programme verknüpft.

Aus den erstellten Kurvengruppen wurde zunächst eine Musterkurve für das Set an N-terminal und C-terminal zugeordneten Kurven ausgewählt. Diesen Musterkurven konnte durch das Vorhandensein vieler definierter Signale aus der 2D-Topologie des Proteins andere Kurven zugeordnet werden. Über die Möglichkeiten einer Visualisierung und Alignmentfunktion des Programmes wurden ausgehend von diesen Anfangsgruppen weitere Kurven in abgestufter Ähnlichkeit nacheinander diesen Mustergruppen zugeteilt. Je nachdem wie einzelne Signale zueinander ähnlich waren, wurden einzelne Kurven mit einbezogen oder wieder ausgeschlossen. Sollten zusätzliche Entfaltungspunkte hinzukommen sein (Abbildung 5.4), wurden diese gekennzeichnet und in der Rubrik „side-Peak“ vermerkt. Peaks, welche anhand der 2D-Topologie und in den Kurvenreihen besonders häufig auftraten, wurden zur Rubrik „Main-Peak“ gezählt.

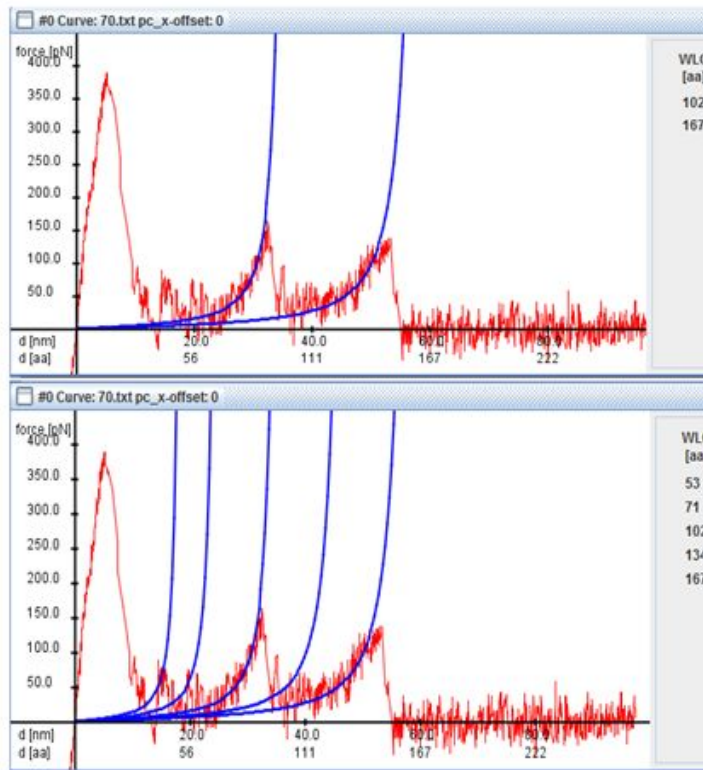


Abbildung 5.4: FD-Kurve der N-terminalen Gruppe mit gekennzeichneten main- und side-Peaks [Cle06]. Visualisiert wurde eine Kurve mit einem typischen N-terminalen Kurvenverlauf. Zu erkennen sind zwei Kraftmaxima mit korrespondierenden Entfaltungslängen von 102 [aa] und 167 [aa]. Nach einer subjektiven Betrachtung wurden drei zusätzliche Signale als mögliche Entfaltungsevents identifiziert (53 [aa], 71 [aa], 134 [aa]) und als side-Peaks bezeichnet.

Der erste Durchlauf berücksichtigte eine maximale Verschiebung der gesamten Kurvengruppen im selbstdurchgeführten Alignment um plus/minus 2 Signale [aa]. Damit das Alignment der einzelnen Kurvengruppen deutlicher wird, wurden die Kurven beim zweiten Durchlauf unabhängig vom oben genannten Grenzbereich, verschoben. Dabei wurden Konturlängen in Bezug auf abgeleitete Längen aus der 2D-Topologie als Bedingung für die Ähnlichkeit festgelegt und die Kurven entsprechend vorhandener Signale passgenau verschoben. Das Ergebnis zeigte ein deutlicheres Bild (Abbildung 5.5).

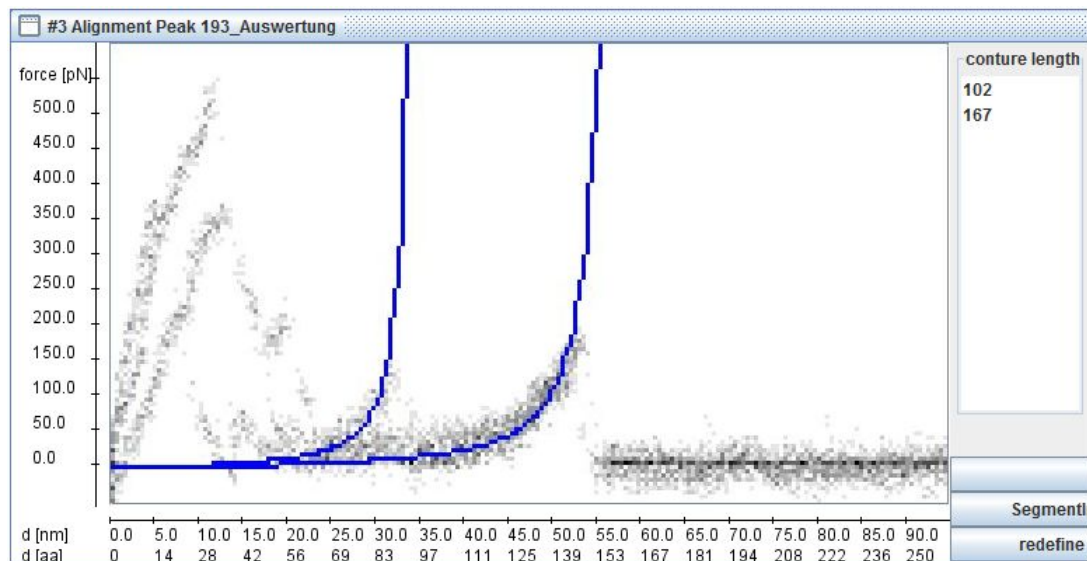


Abbildung 5.5: Kurvenalignment [Cle06]. Das Kurvenalignment zeigt drei alignte Kurven der N-terminalen Gruppe mit einem ähnlichen Entfaltungsweg. Bei allen Kurven konnten Hauptentfaltungslängen von 102 und 167 [aa] ermittelt werden.

Die nötigen Anker und Erkenntnisse aus diesem zweiten Lösungsansatz waren grundlegend und von großer Bedeutung für eine anschließende statistische Analyse einer Verteilung von main- und side-Peaks in den jeweiligen terminalen Gruppen.

5.3 Statistische Analyse vorhandener Entfaltungssignale

Die statistische Analyse des N-terminalen und C-terminalen Datensatzes verfolgte das Ziel, die entsprechenden Entfaltungswege nach dem quantitativen Auftreten vorhandener Konturlängen, zu untersuchen. Weiterführend sollte es im Hinblick auf stabilisierende Bereiche von HP0565 möglich sein, jene statistisch zu validieren. Für die Umsetzung wurden vorwiegend Verteilungen von vorhandenen main-und side-Peaks untersucht und dargestellt.

5.3.1 Verteilung möglicher Entfaltungssignale in Bezug auf den N-terminalen Datensatz

Für die Erstellung der Statistik bezog man sich auf zwei Herangehensweisen. Das erste Konzept berücksichtigte die Informationen aus dem zweiten beschriebenen Lösungsweg. Hauptaugenmerk lag auf allen möglichen Entfaltungslängen [aa], die sich durch die Auswertung des N-terminalen Datensatzes und subjektiven Gesichtspunkten finden ließen. Das Vorkommen aller möglichen Konturlängen wurde mit dem statistischen Merkmal einer relativen Häufigkeit hinterlegt und dargestellt (Abbildung 5.6). In den signifikanten N-terminalen Kurven, konnten in diesem Zusammenhang, Entfaltungsevents in einem Cluster von 32[aa] bis 167 [aa] gefunden werden. Einzelne Entfaltungsevents werden über vorhandene Konturlängen beschrieben.

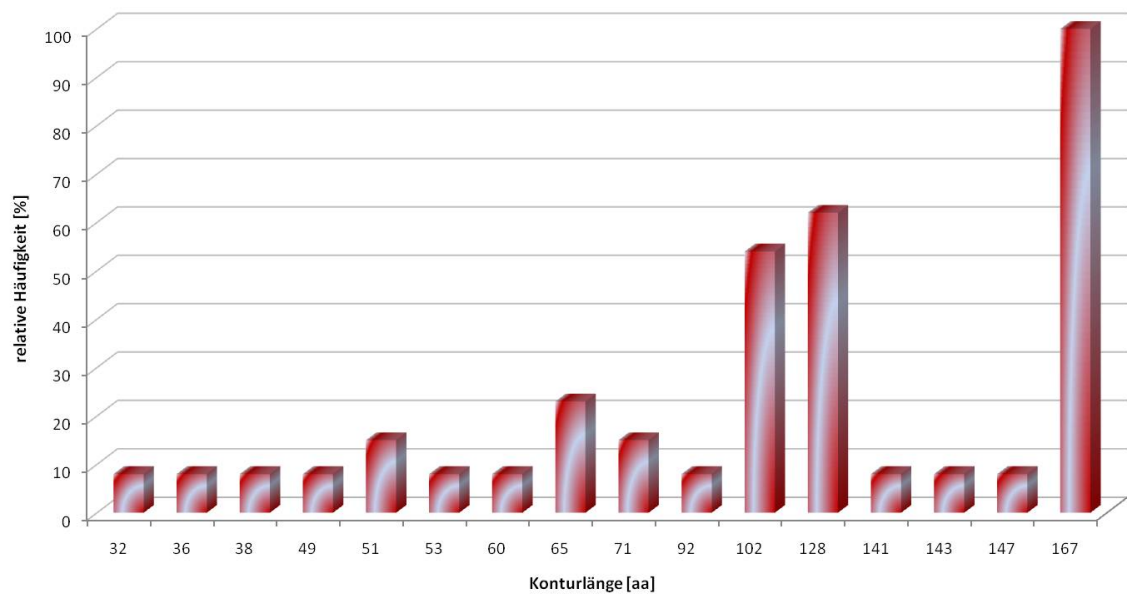


Abbildung 5.6: Verteilung möglicher Entfaltungsevents im N-terminalen Datensatz. Die Darstellung spiegelt das Vorkommen von allen detektierten Konturlängen im N-terminalen Datensatz wider. Auf der x-Achse sind die vorkommenden Längen aufgetragen und die y-Achse beschreibt deren Auftreten im Datensatz über die relative Häufigkeit. Über die Höhen der Balken konnte eine Einteilung in main- und side-Peaks erfolgen.

Deutlich wird, dass die Konturlängen 102 [aa], 128 [aa] und 167 [aa] am häufigsten in der Datenreihe auftreten. Aufgrund des gehäuften Vorkommens, würden diese als main-Peaks bezeichnet werden und definierte Hauptentfaltungswege beschreiben. Somit wäre es möglich, dass das Membranprotein, gezogen vom N-Terminus, über diese Barrieren Haupt-entfaltet. Bei näherer Betrachtung, könnten die membranständigen Bereiche einzeln und nacheinander aus der Membran entfernt wurden sein. Werden lediglich nur die Entfaltungslängen 128 [aa] und 167 [aa] betrachtet, dann wäre es möglich, dass die ersten beiden Helices zusammen entfalten, in einem nachfolgenden Schritt der dritte membranständige Bereich und anknüpfend die vierte Helix. Jedoch liefert die Statistik eine Darstellung über das umfangreiche Vorkommen von vorhandenen side-Peaks, die mögliche Nebenentfaltungswege charakterisieren. Auch diese müssen in einer Stabilitätsanalyse nachfolgend mit betrachtet werden, um Informationen über weitere Stabilitätsbarrieren im Protein abzuleiten. Durch die Verbindung der Statistik mit den Informationen aus der 2D-Topologie können mögliche Entfaltungsevents in den Focus einer Stabilitätsanalyse rücken.

Die zweite und reproduktivere Herangehensweise, berücksichtigte eine Peak-Verteilung, die mit Hilfe von statistischen Funktionen der Analysesoftware [nan10] erstellt wurde. Für die gewünschten Messreihen, in dem Fall der Datensatz von N-terminalen Kurven und C-terminalen Kurven von HP0565, können Peak Statistiken über die Software angefordert werden. Zu den aus den Kurven ermittelten Parametern zählen die minimale sowie eine maximale Entfaltungskraft [pN] und korrelierend eine minimale und maximale auftretende Konturlänge [aa]. In diesem Zusammenhang wird eine Wahrscheinlichkeit des Auftretens der erwähnten Einflussgrößen berechnet und in Prozent wiedergegeben. Aus den Maxima und Minima der beschriebenen Parametern erfolgt eine Mittlung, sodass für jeden detektierten Peak in der Messreihe eine gemittelte Konturlänge und zugehörig die gemittelte Kraft dargestellt werden. Das Zusammenspiel der Parameter kann einen Einblick in die Eigenschaften von Struktursegmenten und deren Wechselwirkungskräften zueinander geben. Diese Erkenntnisse sind wichtig für eine Analyse der Stabilität und Entfaltungsmöglichkeiten des Proteins. Die Software [nan10] arbeitete für diese Untersuchung mit der gleichen Anzahl von FD-Kurven aus den jeweiligen Gruppen, wie bei der vorangegangenen statistischen Betrachtung der Verteilung vorhandener Peaks. Im Gegensatz zur Statistik, die resultierend aus dem zweiten Lösungsweg entstand, werden in der folgenden Darstellung (Abbildung 5.7) die gemittelten Konturlängen [aa] aller existierenden Peaks gezeigt. Die zuvor erstellte Datenreihe stützt sich hauptsächlich auf Konturlängen, welche sich aus der Topologie und der Sequenz ableiten lassen. In Bezug auf die Häufigkeit des Auftretens bestimmter Konturlängen, lassen sich die Längen 119 [aa], 140 [aa], 184 [aa] sichtlich vom Rest vorkommender Entfaltungslängen abgrenzen. Diese würden hier in die Rubrik der „main-Peaks“ zählen und bestimmte Hauptentfaltungswege definieren. Wird das Verschiebungsmaß der Software berücksichtigt und in gemittelter Form abgezogen, so kann ein Vergleich von aufgestellten Entfaltungslängen (Tabelle 5.2) aus der Sequenz von HP0565 und den statistisch ermittelten Entfaltungslängen über diese Analyse erfolgen. Dementsprechend wäre es möglich, dass die oben genannten Konturlängen den Punkten 102 [aa], 128 [aa], 185 [aa] (gemäß der oben verwendeten Reihenfolge) entsprechen. In Anbetracht dieser Informationen, wäre es möglich, dass das Protein in drei aufeinanderfolgenden Schritten hauptentfaltet d.h. erste und zweite Helix entfalten einzeln und im Anschluss die dritte und vierte Helix zusammen, ausgehend vom N-Terminus. Schlussfolgernd lässt sich durch den Vergleich vorgestellter Verteilungen eine Übereinstimmung zwischen zwei der Konturlängen (102 [aa], 128 [aa]) aus beiden statistischen Analysen finden. Auf Grundlage dieser Statistik und in Verbindung mit der 2D-Struktur des Proteins konnte eine Analyse von Entfaltungswegen und eine Stabilitätsanalyse erfolgen.

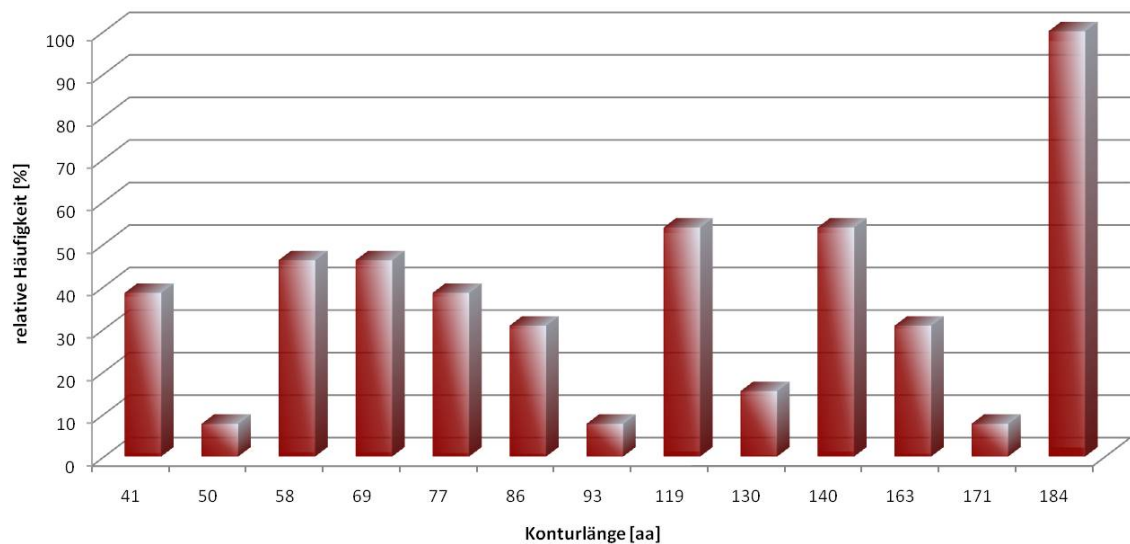


Abbildung 5.7: Konturlängenverteilung im N-terminalen Datensatz (modifiziert nach [nan10]). Dargestellt ist eine über die Software ermittelte Konturlängenverteilung. Auf der x-Achse die detektierten Konturlängen [aa] und über die y-Achse wird erneut das Vorkommen dieser über die Größe einer relativen Häufigkeit dargestellt.

5.3.2 Verteilung möglicher Entfaltungssignale in Bezug auf den C-terminalen Datensatz

Der aus dem zweiten Lösungsweg resultierende C-terminale Datensatz wies eine deutlich geringere Datenmenge im Gegensatz zur N-terminalen Datenreihe auf. Ungeachtet der limitierenden Anzahl von FD-Kurven, wurde auch hier eine Statistik über vorhandene main- und side-Peaks angestrebt. Die Auswertung erfolgte nach dem gleichen Prinzip der Verteilung von Konturlängen aus dem N-terminalen Datensatz (Abschnitt 5.3.1). Die Verteilung aller möglichen Entfaltungslängen der FD-Kurven, die eine C-terminal gezogene Charakteristik aufweisen, zeigt in ähnlicher Weise mögliche Hauptentfaltungsschritte des Proteins. Ebenfalls konnten anfänglich mit Hilfe dieser Informationen, Überlegungen hinsichtlich eventueller Nebenentfaltungsschritte angestellt werden. Aus der Darstellung (Abbildung 5.8) lässt sich erkennen, dass die Signale an der Stelle 89 [aa] und 197 [aa] in der Sequenz von HP0565 verstärkt auftauchen. Bei Beurteilung dieser beiden Entfaltungspunkte, könnte das Protein hauptsächlich über zwei aufeinanderfolgende Schritte entfalten. Analog zur Häufigkeit des Auftretens dieser Konturlängen im Datensatz, schlägt die Statistik vor, dass im ersten Entfaltungsschritt die ersten beiden Helices entfalten und nachfolgend die dritte und vierte Helix (C-terminal gezogen) zusammen in einem Schritt aus der Membran gelöst werden. Auch hier reicht die die Statistik für eine exakte Stabi-

litätsanalyse nicht aus. Eine Verbindung dieser Anhaltspunkte und den Informationen aus der 2D-Struktur von HP0565 ist wesentlich.

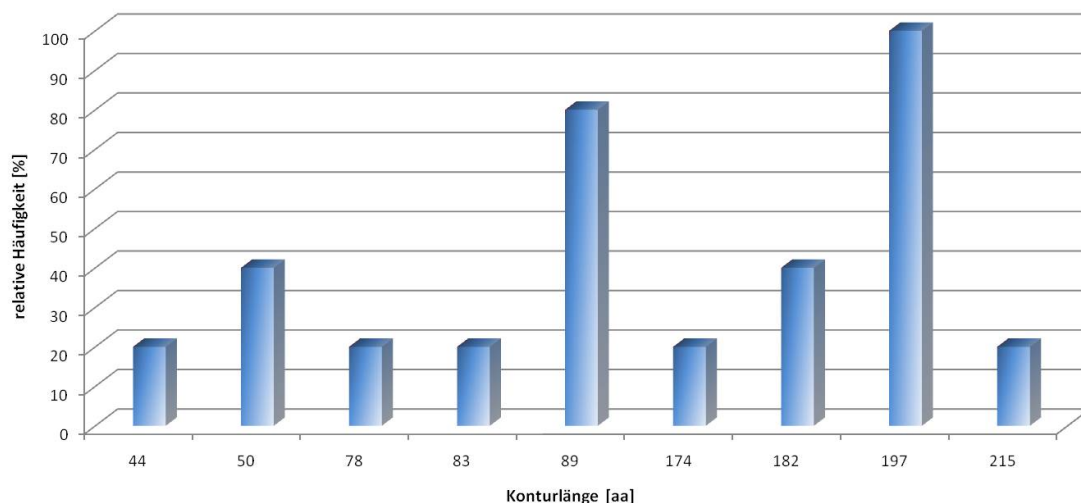


Abbildung 5.8: Verteilung möglicher Entfaltungsevents im C-terminalen Datensatz. Die Darstellung spiegelt das Vorkommen von allen detektierten Konturlängen im C-terminalen Datensatz wider. Auf der x-Achse sind ermittelte Konturlängen aufgetragen und die y-Achse beschreibt deren Auftreten im Datensatz über die relative Häufigkeit. Über die Höhen der Balken konnte eine Einteilung in main- und side-Peaks erfolgen.

Über die statistischen Mittel der Auswertesoftware war es ebenfalls möglich eine Verteilung aller bestehenden Konturlängen [aa] detektierter Peaks in C-terminal beschriebenen FD-Kurven darzustellen (Abbildung 5.9). In Bezug auf die Häufigkeit des Auftretens bestimmter Entfaltungslängen, lassen sich hier die Längen 177 [aa] und 200 [aa] sichtlich vom Rest vorkommender Entfaltungslängen abgrenzen. Im vorderen Bereich der Verteilung treten die Längen 35 [aa], 55 [aa], 67 [aa] gleichhäufig auf. Im Gegensatz zur N-terminalen Datenreihe ist diese Verteilung weitaus unspezifischer und es lassen sich weniger Übereinstimmungen zwischen beiden statistischen Analysen finden. Wird hier in ähnlicher Form vorgegangen und das Verschiebungsmaß der Software berücksichtigt, so kann im gleichen Sinne ein Vergleich von aufgestellten Entfaltungslängen (Tabelle 5.2) aus der Sequenz von HP0565 und den statistisch ermittelten Entfaltungslängen über diese Analyse erfolgen. Werden lediglich nur die Konturlängen 177 [aa] und 200 [aa] betrachtet, würde das bedeuten, dass die Sekundärstrukturelemente des Proteins in einem Schritt aus der Umgebung gelöst werden und das unabhängig von Wechselwirkungsmaxima nur ein spezifisches Kraftmaximum als Entfaltungsweg in der Kraft-Abstandskurve sichtbar wird. Umso wichtiger ist es, dass mit Hilfe dieser Informationen hinführend auf eine Stabilitätsanalyse, insbesondere Nebenentfaltungswege untersucht werden.

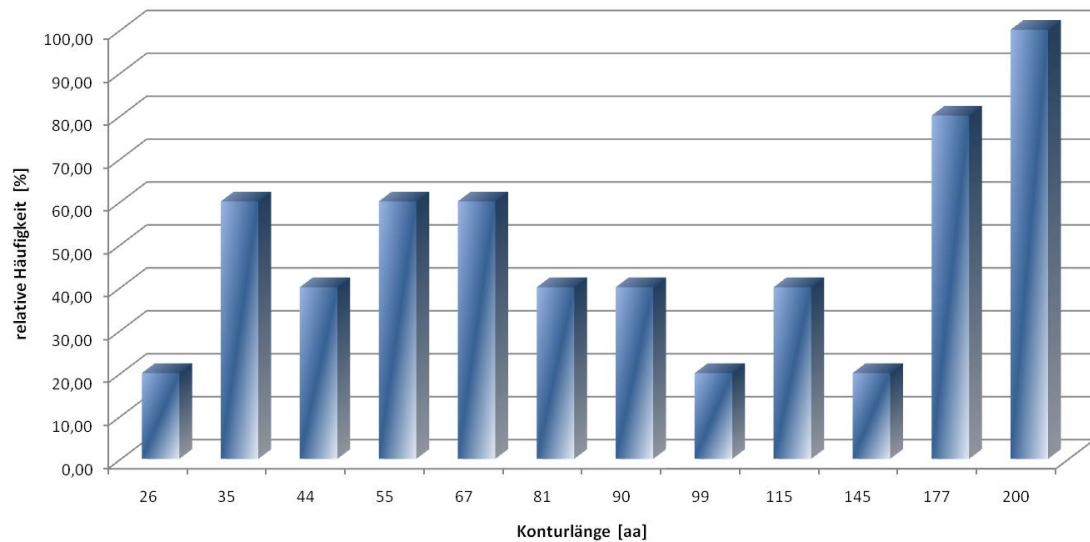


Abbildung 5.9: Konturlängenverteilung im C-terminalen Datensatz (modifiziert nach [nan10]). Dargestellt ist eine über die Software ermittelte Konturlängenverteilung. Auf der x-Achse die detektierten Konturlängen [aa] und über die y-Achse wird erneut das Vorkommen dieser im Datensatz, über die Größe einer relativen Häufigkeit dargestellt.

5.3.3 Verteilung von Kraft und Konturlänge in Bezug auf den N-terminalen Datensatz

Um mögliche Entfaltungsprozesse des Membranproteins HP0565 exakt typisieren zu können, sollte eine Relation zwischen den aufgestellten Konturlängenverteilungen beider Datensätze unter Berücksichtigung einer Verteilung der aufgewendeten Kräfte erfolgen. Die statistischen Kenngrößen der Analysesoftware bildeten für diese Analyse die Basis. Dabei wurden die gemittelten Größen, der über das WLC-Model gefitteten Konturlängen und deren maximale sowie minimale Abweichung, für jeden im Mittelmaß detektierten Peak in beiden Datensätzen dargestellt. Weiterhin erfolgte eine Darstellung der maximalen und minimalen Kraftwerte in Verbindung mit dem gemittelten Wert, so dass sich für jede gemittelte Relation von Kraft und zugehörigem WLC-Fit, eine spezifische Ellipse im Diagramm abzeichnen lies. Dadurch war es möglich Abweichungen beider Parameter in Bezug auf die gemittelten Werte darzustellen und zu interpretieren. Die dargestellte Ziehkraft, kann in diesem Zusammenhang Rückschlüsse auf bestimmte Wechselwirkungsmechanismen (zwischen den Segmenten) zulassen. In Bezug auf den Entfaltungsprozess, kann der Parameter gleichwohl wichtige Informationen über die Anzahl von entfaltenen Strukturelementen liefern. Man geht davon aus, dass mit einer Zunahme von entfaltenen Elementen ein Anstieg im Kraftmaximum erkennbar wird. Dieser Zusammenhang lässt

sich jedoch nur unter Berücksichtigung gleichbleibender Bedingungen beim experimentellen Ablauf übertragen. Diese bedingen mitunter eine gleichbleibende Geschwindigkeit des Zurückfahrens der Tastfeder von der Membran. Unterschiede in den Höhen der Kraftmaxima lassen sich gegebenenfalls durch eine verschiedenartige Lage der Struktursegmente in der entsprechenden Umgebung, im Falle von HP0565 in Mizellenaggregaten, erklären. In der folgenden Darstellung wird die Verteilung von Kraft und Konturlänge bezüglich des N-terminalen Datensatzes aufgezeigt (Abbildung 5.10).

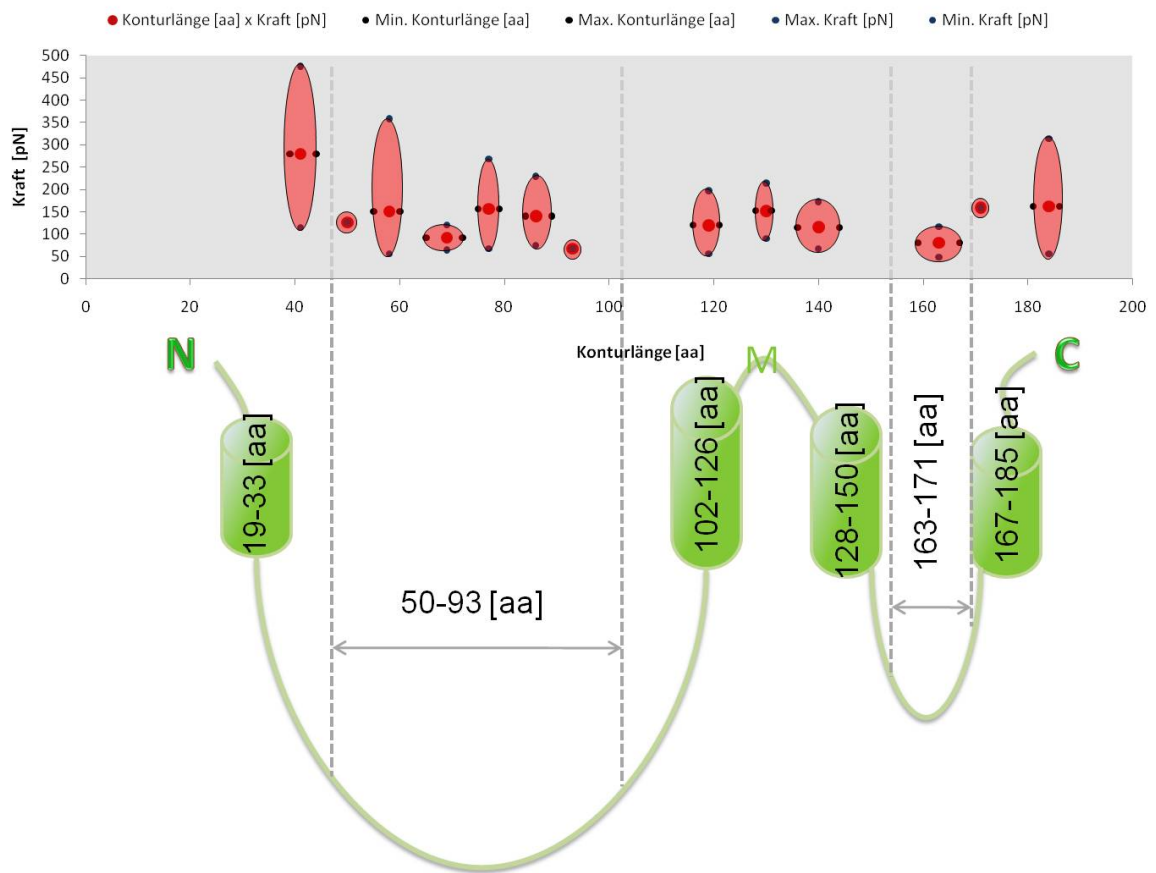


Abbildung 5.10: Verteilung von Kraft und Konturlänge im N-terminalen Datensatz. Die Darstellung vereint die Verteilung der Parameter Kraft und Konturlänge detektierter Peaks des N-terminalen Datensatzes. Eine Längencharakterisierung und Detektion möglicher Entfaltungspunkte spezifischer Sekundärstrukturelemente von HP0565, wurden anhand der Sekundärstruktur vorgenommen (angefangen vom N-Terminus). Im Bereich der langen unspezifischen „Schleife“ wurden sechs Signale gefunden, die einer anschließenden näheren Typisierung bedürfen. Auf der x-Achse aufgetragen: der Konturlängenparameter [aa] und auf der y-Achse wird dieser der Kraftparameter gegenübergestellt

Eine Beurteilung der vorliegenden Verteilung beider Parameter erfolgte nahe der festgelegten Entfaltungslängen aus der 2D-Topologie, um strukturelle Entfaltungspositionen in die Analyse einzubeziehen. Anhand der Größe der Ellipsen war es möglich die jeweiligen Abweichungen des betrachteten Parameters für den betrachteten Peak abzuleiten und somit bedeutende Informationen über den Entfaltungsprozess zu erhalten. Die größte Differenz von minimal und maximal aufgetragener Kraft wird bei den Konturlängen 41 [aa], 58 [aa] und 184 [aa] deutlich. An diesen Positionen variiert die durchschnittliche Kraft um mehr als 200 [pN]. Diese Abweichung lässt sich möglicherweise durch ein verschiedenartige Neigung der Strukturelemente zueinander erklären. Eine nichtsymmetrische Lage der einzelnen Segmente verlangt eine erhöhte Entfaltungskraft wenn die Bereiche mechanisch aus der Membran entfernt werden. Eine hohe Konturlängenabweichung lässt sich an den sequenziellen Positionen 69 [aa], 140 [aa] und 163 [aa] feststellen. Anhand der Topologie von HP0565 lässt sich die Position 69 [aa] der unspezifischen Loop Region zuweisen, jedoch keinem membranständigen Bereich. Die Abweichung an Stelle 140 [aa] kann eine mögliche Entfaltungsbarriere in der Mitte des dritten transmembranen Bereiches beschreiben. In Bezug auf die vorhandene Abweichung wird hier die Barriere in der Sequenz um ein Längenmaß von vier Aminosäuren verschoben. Diese Entfernung ruft jedoch keine Positionsänderung in Richtung extra-oder intrazelluläres Ende des transmembranen Bereiches hervor. Die Schwankung an Stelle 163 [aa] bezieht sich auf das intrazelluläre Ende des zweiten outside Bereiches. Mit einer Fluktuation von maximal drei Aminosäuren, kommt es zu keiner signifikanten Verlegung dieser Barriere. Allerdings charakterisiert diese Barriere keinen transmembranen Bereich. Durch die Verknüpfung von Topologie und Verteilung, wird eine Lage mehrerer Signale in dem unspezifischen coil Bereich (50-93 [aa]) in der 2D-Struktur erkennbar. Daher galt es diesen Bereich noch einmal genauer, anhand einer transmembran gestützten Motivsuche zu spezifizieren, indem mögliche Helixcaps detektiert werden.

5.3.4 Verteilung von Kraft und Konturlänge in Bezug auf den C-terminalen Datensatz

Eine Einschätzung, der in die Verteilung einfließenden Parameter, erfolgte in dem selben Sinne wie die Beurteilung des N-terminalen Datensatzes (Abbildung 5.11). Anhand der Beschaffenheit der dargestellten Ellipsen, war es auch hier möglich betreffende Abweichungen zu diskutieren. Die größte Differenz von minimal und maximal aufgetragener Kraft wird bei den Mittelpunkten 35 [aa], 55 [aa] und 67 [aa] deutlich. An diesen Positionen schwankt die durchschnittliche Kraft ebenfalls um mehr als 200 [pN]. Dieses Phänomen kann gleichermaßen durch eine veränderliche Position der transmembranen Elemente bedingt sein. Eine erhöhte Konturlängenabweichung lässt sich allein an gemittelter Position 177 [aa] erkennen. Die Streuung an anderen Positionen ist auffällig gering. Durch ein Hinzuziehen der Topologie von HP0565 lässt sich die Position 177 [aa] der unspezifischen Loop Region zuweisen, jedoch ebenfalls keinem membranständigen Bereich.

Diesen Hintergrund galt es durch die genannte Analyse von möglichen Helixcap-Motiven zu verstehen. Im Vergleich zur Darstellung des N-terminalen Datensatzes ist in der Verteilung der C-terminalen Datenreihe eine weitaus geringere Abweichung der Parameter deutlich. Jedoch umfasst dieser Datensatz bedeutend weniger Daten als der N-terminale, wodurch eine Reproduzierbarkeit an dieser Stelle angezweifelt werden kann.

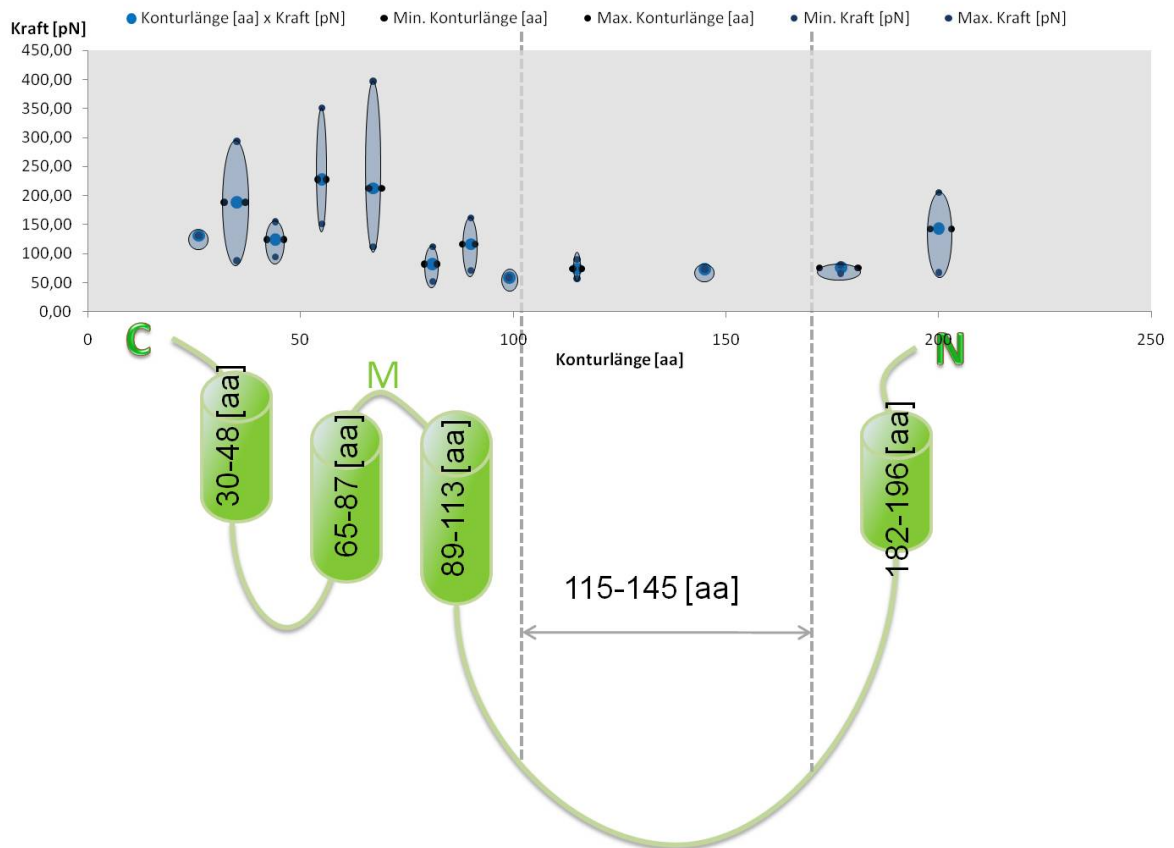


Abbildung 5.11: Verteilung von Kraft und Konturlänge im C-terminalen Datensatz. Die Darstellung vereint die Verteilung der Parameter Kraft und Konturlänge detektierter Peaks. Eine Längencharakterisierung und mögliche Feststellung von Entfaltungspunkten spezifischer Sekundärstrukturelemente von HP0565 wurden anhand der Sekundärstruktur vorgenommen, angefangen vom C-Terminus. Im Bereich der langen unspezifischen „Schleife“ wurden zwei Signale gefunden, die einer anschließenden näheren Typisierung bedürfen. Auf der x-Achse aufgetragen: der Konturlängenparameter [aa] und auf der y-Achse wird dieser der Kraftparameter gegenübergestellt.

5.4 Spezifizierung von Stabilitätsbarrieren (Stabilitätsanalyse)

Die meisten Membranproteine entfalten häufig nicht nur über einen definierten Entfaltungsweg. In den Messreihen, die über SMFS aufgenommen werden, sind in vielen Fällen divergente Entfaltungswege nachweisbar. Durch mögliche Konformationsänderungen des Proteins in der Membran können parameterbedingte Änderungen im Kraftspektrum auftauchen. Eine Verknüpfung der im Vorfeld beschriebenen Resultate aus der statistischen Analyse, insbesondere von identifizierten main- und side-Peaks in den Datensätzen und den bedeutenden Informationen aus der 2D-Topologie von HP0565, erlauben

ein strukturelles mapping von Stabilitätsbarrieren. Somit können durch diese Festlegung, Haupt- und Nebenentfaltungswege des Proteins spezifiziert werden. Für diese umfassende Betrachtung waren die bewerteten Konturlängenverteilungen beider Datenreihen mit ihrer Häufigkeit, die über die Peakverteilung der Analysesoftware [nan10] entstanden, grundlegend. Folgende Abbildungen verdeutlichen mögliche Entfaltungswege für den N-terminalen und C-terminalen Datensatz. Für diese Darstellung wurden drei signifikante Kurven ausgewählt. Jede der aufgenommenen FD-Kurven kennzeichnet einen eventuellen Entfaltungsweg über markierte main- und side-Peaks. Die anschließenden Aussagen wurden unter Berücksichtigung gleichbleibender experimenteller Bedingungen getroffen. Das sich ein Kurvenbild aufgrund einer Änderung des pH-Wertes hätte verändern können, wurde hierbei ausgeschlossen. Die Abbildung 5.13 gibt Aufschluss über die Reihenfolge, in der die transmembranen Bereiche entfalten und ob diese einzeln oder in Kombination mit anschließenden Elementen entfernt worden sind.

Das obere Kurvenmuster (I) (Abbildung 5.12) kennzeichnet einen ersten möglichen Entfaltungsweg, wenn das Protein mit dem N-terminalen Ende aus der Umgebung entfernt wurde. Dabei kommt es zu einer kompletten Entfaltung des ersten transmembranen Bereiches in Richtung extrazelluläres Ende der ersten Helix (WLC = 41 [aa]). Durch Anwendung einer Kraft von ca. 115 [pN] ist ein nächster Entfaltungsschritt bis zur Hälfte der zweiten Helix (WLC = 119 [aa]) möglich. Bei Überschreiten einer entsprechenden Stabilität wird die dritte N-terminale Helix komplett gelöst und eine weitere Barriere lässt sich wiederholt am extrazellulären Ende des vierten helikalen Bereiches feststellen, sodass durch diesen Schritt die Länge des entfalteten Polypeptids auf 171 [aa] ansteigt. Im vierten und letzten Entfaltungsevent lässt sich ein Signal beschreiben, welches durch die Nähe zum cytoplasmatischen Ende des vierten membranständigen Bereiches lokalisiert ist und durch lösen dieses Komplexes aus der Umgebung zustande kommt (184 [aa]). Unter Berücksichtigung des Kurvenverlaufs, der im oberen Bereich der Abbildung dargestellten FD-Kurve, ist davon auszugehen, dass das Maximum bei 171 [aa] eher unspezifisch einzuordnen ist und folglich das Maximum bei 184 [aa] die letzten beiden Entfaltungsschritte einschließt, sodass die Barriere bei 119 [aa] den letzten möglichen Entfaltungsvorgang kennzeichnet. Demzufolge würde ein letzter Abschnitt von 65 Aminosäuren der Polypeptidkette gestreckt werden. Werden die Wahrscheinlichkeiten der genannten Konturlängen in die Beurteilung integriert, so kann hervorgehoben werden, dass die Konturlänge 184 [aa] mit einer Wahrscheinlichkeit von 100 % und der Entfaltungspunkt bei 119 [aa] zu 53 % auftritt. Infolgedessen ist es sehr wahrscheinlich, dass das Protein bei diesem speziellen Entfaltungsprozess über die genannten Barrieren entfaltet. Die Konturlängen 41 [aa] und 171 [aa] treten sehr selten im Datensatz auf. Aufgrund dessen würde diesen möglichen Entfaltungspunkten eine eher unspezifische Position zugeordnet werden.

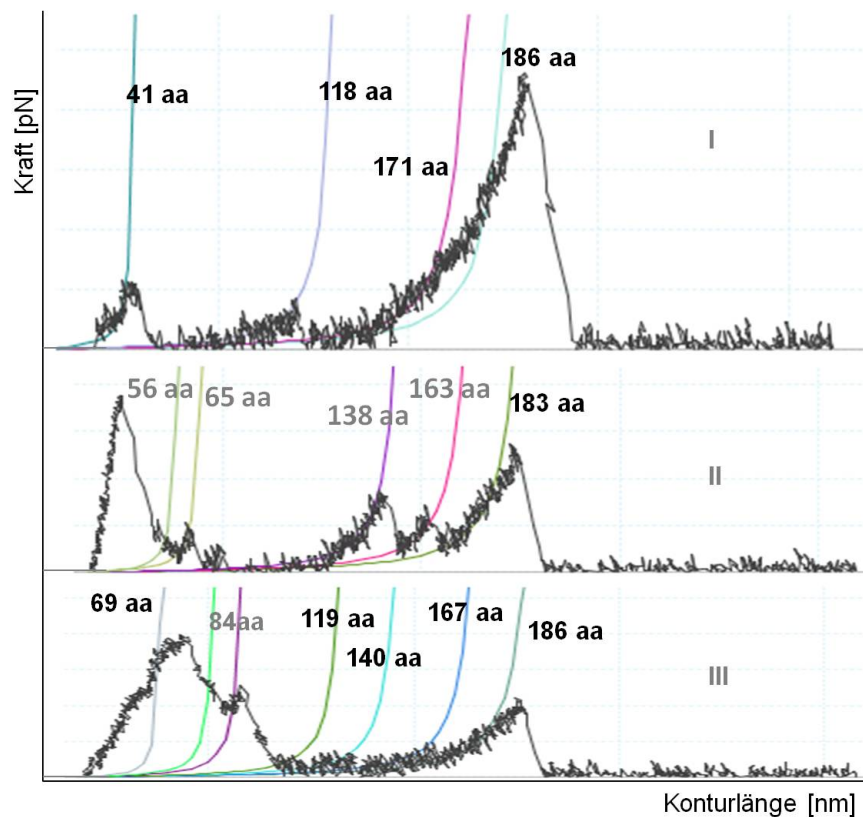


Abbildung 5.12: Mögliche Entfaltungswege für den N-terminalen Datensatz (modifiziert nach [nan10]). In der Abbildung sind durch die Software tatsächlich ermittelten Konturlängen der Entfaltungsevents, vorgestellter FD-Kurven, dargestellt. Auf der y-Achse ist die Kraft als Richtparameter aufgetragen. Grau hinterlegte Konturlängen kennzeichnen die in den Entfaltungswegen hinzukommenden Konturlängen. In Bezug auf die Charakterisierung spezifischer Entfaltungswege bezog man sich jedoch auf die gemittelten Werte vorhandener Entfaltungslängen im Datensatz der N-terminalen Reihe.

Das zweiten abgebildete Kurvenmuster (II) stellt einen weiteren signifikanten Entfaltungsprozess von HP0565 dar. Durch hinzukommende Konturlängen werden hier weitere Stabilitätsbarrieren vorgestellt und spezifiziert. Eine Entfaltung über die Barrieren 56 [aa] und 65 [aa] kennzeichnet das Lösen der ersten Helix bis in den anschließenden unspezifischen Loop (outside) Bereich. Ein anknüpfender Schritt beschreibt die Entfaltung des kompletten zweiten helikalen Bereiches über den folgenden inside-Bereich, einschließlich des Anfangsbereiches der dritten Helix. Zu diesem Zeitpunkt umfasst die Streckung der Polypeptidkette 140 (140 [aa]) Aminosäuren. Weiteres Zurückziehen der Tastspitze dehnt das Polypeptid weiter (163 [aa]) und die Kraft greift unmittelbar am bisher noch gefalteten Bereich der dritten transmembranen Region an, sodass sich diese (Abbildung 5.13) über die anhängende cytoplasmatischen Schleife entfaltet. Ähnlich wie bei dem zuerst beschriebenen Entfaltungsweg schließt sich diesem Prozess eine Entfaltung über

die Position 184 [aa] an. Unter Einbeziehung der Auftrittswahrscheinlichkeiten genannter Peaks im Datensatz, tauchen die Positionen 56 [aa], 65 [aa], 140 [aa] und 184 [aa] am häufigsten auf. Diese main-peaks deuten im beschriebenen Entfaltungsweg auf signifikante Hauptentfaltungspunkte hin.

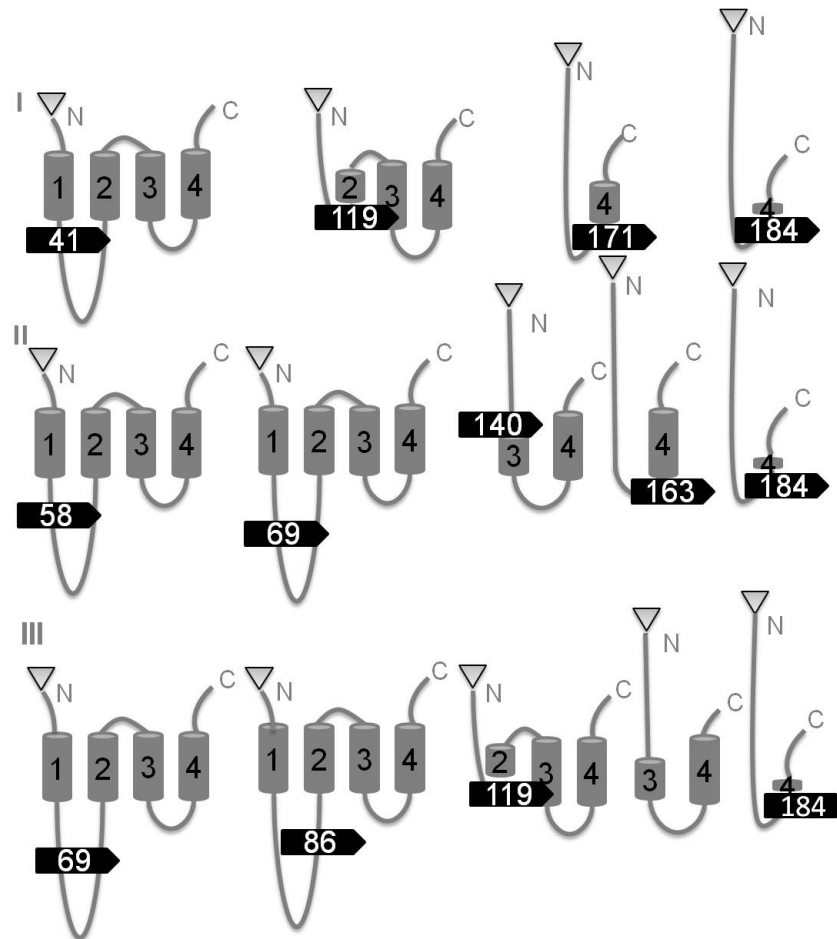


Abbildung 5.13: Mögliche Entfaltungswege für den N-terminalen Datensatz anhand der Topologie von HP0565. Die Abbildung zeigt identifizierte Entfaltungsbarrieren (schwarz hinterlegt) in der 2D-Struktur von HP0565, entsprechend des jeweiligen Entfaltungsweges (I-III). Die Position schwarz markierter Entfaltungsbarrieren richtete sich nach den gemittelten Werten vorhandenen Konturlängen aus der Statistik. Beurteilt wurden jedoch tatsächlich detektierte Peaks mit deren Auftrittswahrscheinlichkeiten. Die transmembranen Helices wurden ausgehend vom N-Terminus aufsteigend nummeriert.

Die dritte FD-Kurve aus dem N-terminalen Datensatz von HP0565 zeigt einen Entfaltungsweg (III), welcher schon beschriebene main-und side-Peaks sowie bestimmte Entfaltungsschritte vorangegangener Muster einschließt. Ein zusätzlicher für diese Kurve signifikanter Entfaltungsschritt wird über die Barriere bei 84 [aa] aufgezeigt, die allerdings mit einer Wahrscheinlichkeit von knapp 30% weniger auftritt.

Die Identifizierung möglicher Entfaltungsbarrieren für den C-terminalen Datensatz erfolgte nach dem gleichen Prinzip wie die vorgenommene Stabilitätsbetrachtung der N-terminalen Datenreihe. Das Prinzip basierte ebenfalls auf Informationen aus den ermittelten Entfaltungslängen und den statistischen Kenngrößen. Das erste spezifische Kurvenmuster (I) (Abbildung 5.14) deutet einen ersten möglichen Entfaltungsweg an, wenn das Protein mit dem C-terminalen Ende aus der Umgebung entfernt wird. Indessen wird bei Überwinden der Barriere 44 [aa] ein Entfalten des Anfangsbereiches der ersten C-terminalen Helix vorgeschlagen. Im zweiten Schritt ist eine Entfaltung des noch vorhandenen membranständigen Bereiches möglich, wobei auch hier die entfaltete Region bis in den ersten anschließenden outside Bereich (56 [aa]) reicht. Dies wird anhand der 2D-Topologie erkenntlich (Abbildung 5.15).

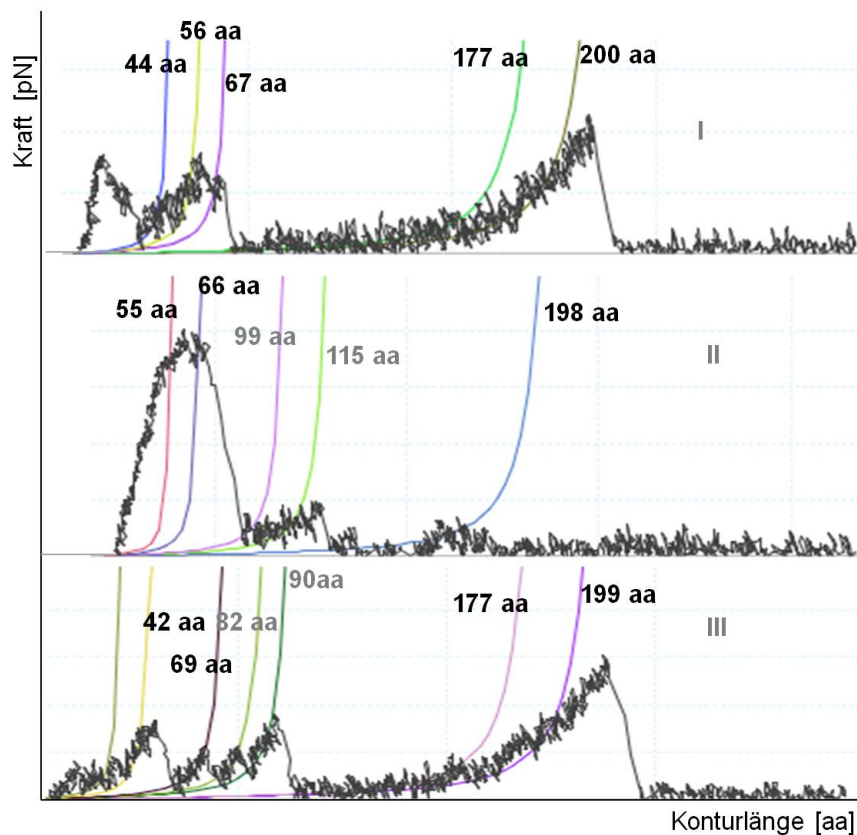


Abbildung 5.14: Mögliche Entfaltungswege für den C-terminalen Datensatz (modifiziert nach [nan10]). In der Abbildung sind durch die Software tatsächlich ermittelten Maxima der Entfaltungsevents vorgestellter FD-Kurven dargestellt, in Verbindung mit der entsprechenden Konturlänge [aa]. Auf der y-Achse ist die Kraft als Richtparameter aufgetragen. Grau hinterlegte Konturlängen kennzeichnen die in den Entfaltungsweegen hinzukommenden Konturlängen. In Bezug auf die Charakterisierung spezifischer Entfaltungswege bezog man sich jedoch auf die gemittelten Werte vorhandener Entfaltungslängen aus der Peak-Statistik des Datensatzes der C-terminalen Reihe.

Somit wäre es möglich, dass der erste membranständige Bereich in die intrazelluläre Region aus der Membran heraus ragt. Nachfolgend findet sich eine Barriere zum zytoplasmatischen Areal (67 [aa]). Im dargestellten Kurvenverlauf ist dieser Entfaltungspunkt sehr unspezifisch und charakterisiert keinen eigenständigen Entfaltungsschritt. Viel mehr wäre es denkbar, dass dieser den vorangegangenen Entfaltungsschritt beinhaltet. Anschließend würden der verbleibende Bereich der zweiten Helix zusammen mit der gesamten dritten helikalen Region bis zur intrazellulären Region der vierten C-terminalen Helix entfalten.

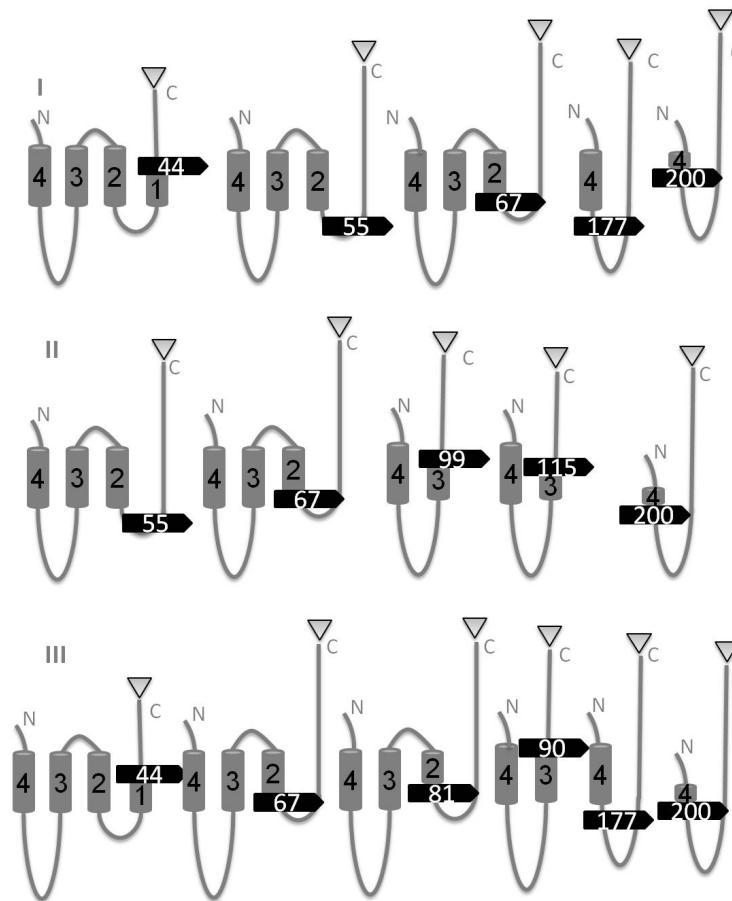


Abbildung 5.15: Mögliche Entfaltungswege für den C-terminalen Datensatz anhand der Topologie von HP0565. Die Abbildung zeigt identifizierte Entfaltungsbarrieren (schwarz hinterlegt) in der 2D-Struktur von HP0565, entsprechend des jeweiligen Entfaltungsweges (I-III), wenn das Protein mit dem C-Terminus an die Tastspitze adsorbiert. Die Position schwarz markierter Entfaltungsbarrieren richtete sich nach den gemittelten Werten vorhandener Entfaltungslängen aus der Statistik. Beurteilt wurden jedoch tatsächlich detektierte Peaks mit deren Aufttrittswahrscheinlichkeiten. Die transmembranen Helices wurden ausgehend vom C-Terminus aufsteigend nummeriert.

Unter Einbeziehung der Wahrscheinlichkeiten werden die Konturlängen 177 [aa] und 200 [aa] zu Hauptentfaltungsbarrieren. Die gestreckte Kette von 200 [aa] ist durchweg in allen festgelegten FD-Kurven mit einem C-terminalen Kurvenverlauf zu finden. Mit einem Vorkommen von 80 % ist die Konturlänge 144 [aa] sehr häufig in der Datenreihe vertreten. In der zweiten vorgestellten FD-Kurve (II) werden erläuterte Entfaltungslängen (56 [aa], 66 [aa], 198 [aa]) durch Barrieren an den Stellen 99 [aa] und 115 [aa] ergänzt. Wobei es auch hier möglich wäre, dass die Konturlänge bei 115 [aa] den detektierten möglichen Entfaltungsschritt einbezieht. Die dritte FD-Kurve aus dem C-terminalen Datensatz von HP0565 zeigt einen Entfaltungsprozess (III), welcher ebenso beschriebene main- und side-Peaks sowie bestimmte Entfaltungsschritte vorangegangener Wege einschließt. Zu-

sätzliche, für diese Kurve signifikante Entfaltungsschritte, werden über die Stabilitätsbarriere bei 82 [aa] und 90 [aa] aufgezeigt, die allerdings mit einer Wahrscheinlichkeit von knapp 30% weniger auftritt.

Anhand der Aussagen aus der vorangegangenen Analyse nach spezifischen Stabilitätsbarrieren und Entfaltungswegen des Proteins, war eine weitere Motivsuche nach bestimmten strukturellen Präferenzen bzw. Domänen notwendig. In diesem Fall bezog sich die Motivsuche jedoch nicht auf ein Identifizieren von funktionellen Bereichen. Die Untersuchung erbrachte eine Überprüfung bereits vorhergesagter Helices aufgrund von unspezifischen Entfaltungsbarrieren in inside und outside Regionen der Topologie von HP0565. N-terminal spielten dabei folgende, in den FD-Kurven auftretende Konturlängen, eine Rolle: 41 [aa], 58 [aa], 69 [aa], 86 [aa], 163 [aa]. Die erwähnten Längen kennzeichnen mit Ausnahme der Länge 163 [aa], Barrieren, die innerhalb der ersten längeren Loop-Region (34-101[aa]) positioniert sind und keinen transmembranen Bereich signalisieren. Im Voraus wurde diese Region gesondert auf weitere Sekundärstrukturelemente geprüft. Die Ergebnisse zeigten dass sich möglicherweise Faltblatt- bzw. hinzukommende Helixstrukturen dort befinden könnten. Eine sichere Vorhersage konnte nicht erzielt werden. Aus diesem Grund standen bereits vorhergesagte Helix-Domänen für die Untersuchung im Vordergrund. Eine Assoziation von Sekundärstrukturelementen von HP0565, Primärstruktur sowie einer Abfolge der über die Peak-Verteilung ermittelten möglichen Entfaltungsbarrieren, verdeutlicht den Sachverhalt für die N-terminale Betrachtung (Abbildung 5.16).

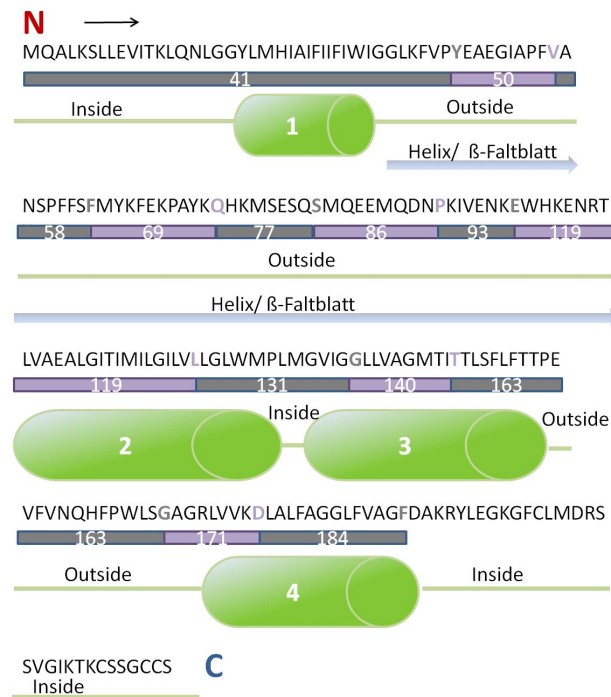


Abbildung 5.16: Assoziation von Topologie, Sequenz und Entfaltungsbarrieren von HP0565 (N-terminal). Die Darstellung vereint eine Abfolge vorhergesagter Sekundärstrukturelemente (grün) und deren sequenzielle Position. Abwechselnd werden mögliche Barrieren (lila, grau), welche sich durch die entsprechende Peakverteilung ergeben haben, aufgezeigt. Genannte Barrieren sind durch die jeweilige Aminosäure, ausgehend von den Entfaltungspunkten, farblich gekennzeichnet (lila, grau). Der Pfeil signalisiert das Entfernen der Proteinkette ausgehend vom N-Terminus. Die erste lange Schleife wird in dieser Darstellung mit einem blauen Pfeil gekennzeichnet. Durch die Analyse könnten dort weitere bzw. verlängerte Helix- oder Faltblattstrukturen lokalisiert sein. Die Abbildung bildet eine Grundlage für ein lokalisieren eventueller Bereichserweiterungen.

C-terminal wurden die Maxima an Stelle 55 [aa] und 177 [aa] näher untersucht (Abbildung 5.17). Die Positionen genannter Barrieren im Hinblick auf die 2D-Struktur des Proteins, deuten häufig auf eine Lage nahe den intrazellulären sowie extrazellulären Helixenden bzw. Helixanfängen hin, je nachdem welches terminale Ende betrachtet wird, so dass an dieser Stelle eine Überprüfung auf mögliche Erweiterungen von membranständigen Bereichen im Mittelpunkt stand.

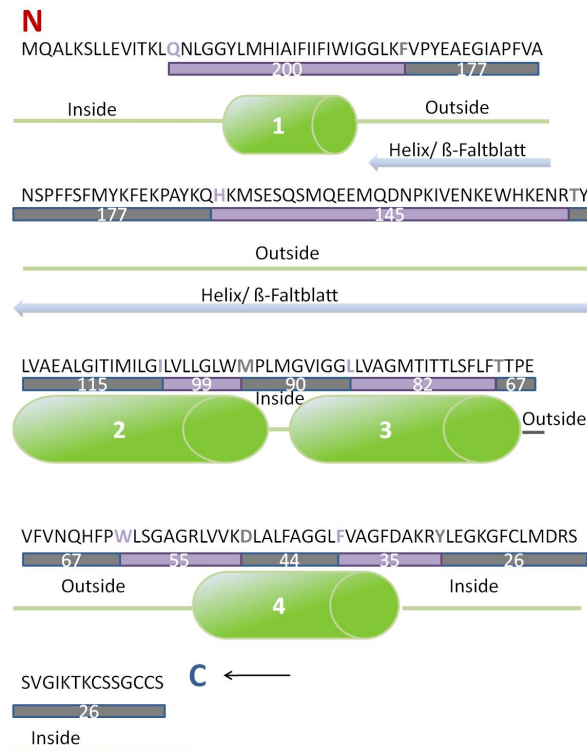


Abbildung 5.17: Assoziation von Topologie, Sequenz und Entfaltungsbarrrieren von HP0565 (C-terminal). Die Darstellung vereint eine Abfolge vorhergesagter Sekundärstrukturelemente (grün) und deren sequenzielle Position. Abwechselnd werden mögliche Barrieren (lila, grau), welche sich durch die entsprechende Peakverteilung ergeben haben, aufgezeigt. Genannte Barrieren sind durch die jeweilige Aminosäure, ausgehend von den Entfaltungspunkten, farblich gekennzeichnet (lila, grau). Die lange Schleife wird in dieser Darstellung mit einem blauen Pfeil gekennzeichnet. Durch die Analyse könnten dort weitere bzw. verlängerte Helix- oder Faltblatt strukturen lokalisiert sein. Der Pfeil signalisiert das Entfernen der Proteinkette ausgehend vom C-Terminus. Die Abbildung bildet eine Grundlage für ein lokalisieren eventueller Bereichserweiterungen.

Eine besondere Form von erweiterten konservierten Helixregionen stellen die Helixcap-motive dar. Die Helixcaps sind folglich Motive zur Stabilisierung des Übergangsbereiches von einer Helix zur benachbarten Sekundärstruktur. Als transmembran wird der Teil bezeichnet, der sich ausschließlich im hydrophoben Milieu der Membran befindet. Caps ragen aus der Membran heraus, zählen jedoch zum helikalen Bereich. Unter Zuhilfenahme von Aminosäurepräferenzen für N- und C-Cap Positionen von Turn-Helix Motiven wurden für jede der vier Helices Möglichkeiten gesucht, weitere Aminosäuren den Bereichen anzuschließen. Eine Suche nach bereits annotierten Helixcap-Motiven lieferte keine Informationen. Um die Tendenzen der Aminosäuren bestimmte Sekundärstrukturelemente auszubilden, wie alpha-Helix und beta-Faltblatt, zu terminieren, wurden zunächst die Chou Fasman Parameter einbezogen. Dadurch war es möglich die helikalen Sequenzabschnitte

und einzelne dazugehörige Aminosäuren nach diesen Präferenzen zu bewerten. Die Chou Fasman Parameter beabsichtigen eine Zuordnung der 20 proteinogenen Aminosäuren in folgende Kategorien: Starke Helixbildner (Glu, Ala, Leu), Helixbildner (His, Met, Gln, Trp, Val, Phe), Schwache Helixbildner (Lys, Ile), Indifferente Aminosäuren (Asp, Thr, Ser, Arg, Cys), Schwache Helixbrecher (Asn, Tyr), Starke Helixbrecher (Pro, Gly). Diese Klassifizierung gilt analog für eine Zuordnung in die Sekundärstruktur beta-Faltblatt und lediglich für globuläre Proteine. Die Untersuchung der Chou Fasman Parameter der Aminosäuren in Abhängigkeit der Lage im Protein (außerhalb der Membran oder Transmembran) zeigte überraschenderweise, dass die Chou Fasman Parameter für Bereiche außerhalb der Membran keine Tendenzen zu den bekannten Parametern in globulären Proteinen zeigten. Daraus lässt sich ableiten, dass die chemische Zusammensetzung in outside Bereichen stark von den globulären Proteinen unterscheidet. Hauptsächlich wurden spezifische Eigenschaften einzelner Aminosäuren für eine Abschätzung einbezogen. In diesem Zusammenhang wurden Einteilungskriterien, ob sich die betrachtete Aminosäure bevorzugt in Anfangs- bzw. Endbereichen befindet sowohl bei N-als auch bei C-Terminus, ob diese im Helixkern lokalisiert ist oder eher in Regionen außerhalb dieser Abschnitte und ob diese in Bezug auf deren physiko-chemischen Eigenschaften einen hohen amphiphilen Charakter aufweisen und somit häufig im Übergang von hydrophilem zu hydrophobem Milieu auftreten. In Bezug auf Helix Cap Motive konnte eine Einteilung nach spezifischen N-Cap und C-Cap Positionen vorgenommen werden. Durch jene Zuweisung der Aminosäuren wurde jeder bereits vorhergesagte transmembrane Abschnitt mit zugehörigen Aminosäuren der vorangegangenen bzw. nachfolgenden inside/ outside Regionen untersucht. Die folgende Abbildung (Abbildung 5.18) einzelner sequenzieller Helices-Abschnitte geben Aufschluss über Resultate dieser Überlegung. **(I)** W und I könnten Cap-Position C-terminal einnehmen, wobei W aufgrund des amphiphilen Charakters auch eine Übergangsposition aus der TM-Helix in die hydrophile Umgebung, C-terminal, charakterisiert. Y tritt im Zusammenhang mit G im Übergang von einem hydrophilen in ein hydrophobes Milieu der TM-Helix auf. Ein typisches Muster was den Helix-Kern beschreibt wird durch eine Abfolge der Aminosäuren M-H-I-A-I-F-I-I-F-I vorgeschlagen. I, A, M meist im Helixkern lokalisiert. Unter Einbeziehung von jeweils acht Aminosäuren am Anfang (inside-Bereich) und Ende (outside-Bereich) des Abschnittes, könnte der TM-Bereich um eine Position am Anfang verlängert werden. G ist oft außerhalb der TM-Helix zu finden (Position 34 und 35 in der Sequenz). **(II)** Motiv L-V-A-E-A ist N-terminal als Cap möglich. Auch innerhalb der Helix wäre dieses denkbar. Nachfolgende Aminosäuren können den Membrankern signalisieren. **(III)** Das Motiv G-V-I-G-G kann die Helix einleiten. T-I-T-T bildet ein präzises Paket an Aminosäuren, welche sich charakteristisch innerhalb der Helix befinden. F-T-T-P kennzeichnet den Übergang von hydrophob

in hydrophil. Prolin könnte sich an den Helixkern anschließen. **(IV)** L-V-V-K-D ist als Motiv für eine N-Cap Position möglich. Die Abfolge der Aminosäuren L-A-L-F-A-G-G-L-F wäre signifikant für das Innere der Helix. Die Abfolge G-A-G-R könnte eine Verlängerung am Anfang der Helix darstellen oder zum angedeutenden N-Cap. Sie sind eher untypisch außerhalb einer TM-Helix. Die Abfolge D-A-K-R signalisiert möglicherweise eine C-cap-Charakteristik.

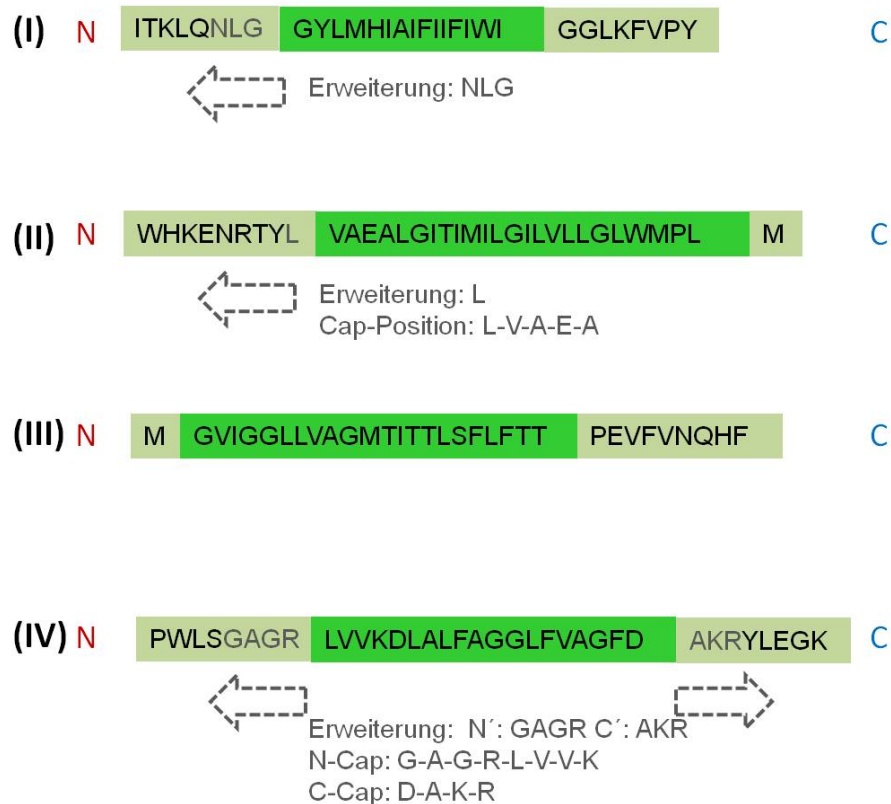


Abbildung 5.18: Motivsuche in vorhergesagten transmembranen Regionen von HP0565. Transmembrane Bereiche sind dunkelgrün hinterlegt. Einzelne Aminosäuren der insode-und outside-Regionen hellgrün. Grau hinterlegte Aminosäuren deuten eine mögliche Verlängerung in Richtung C- bzw. N-Terminus oder eventuelle N- bzw. C-Cap Motive an. Die Nummern (I) bis (IV) verweisen auf die vier vorhergesagten Helices des Proteins. Rot sind N-terminale Enden und blau, C-terminale Enden hervorgehoben..

Mit Hilfe dieser Anhaltspunkte ist eine theoretische Erweiterung bereits vorhergesagter membranständiger Regionen möglich. Über physiko-chemische Eigenschaften einzelner Aminosäuren und weiteren Präferenzen lässt sich topologisch eine Veränderung der ersten helikalen Struktur mit einem Sequenzbereich von jetzt an 16 [aa] bis 33 [aa] und für die vierte Helix ein Bereich von 163 [aa] bis 188 [aa] abgrenzen. Diese Angaben richten sich nach einer N-terminalen Entfaltungscharakteristik, entsprechend vorgestellter möglicher Entfaltungswege und den Informationen aus der Statistischen Analyse. Zwar konnte das Maximum an Stelle 163 [aa] lediglich der Kategorie „side-Peaks“ zugeteilt werden, jedoch kann dieses Maximum an dieser Stelle durch die helikale Ausdehnung erklärt werden. C-terminal lässt sich durch diese Informationen die ermittelte Barriere an Position 52 [aa] einer weiteren möglichen Entfaltung einer helikalen Struktur zuordnen. Allerdings ist der Informationsgehalt aus dieser Analyse nicht maßgeblich.

Kapitel 6

Ausblick

Es konnte gezeigt werden, dass es möglich ist Membranproteine an Mizellen mittels SMFS zu analysieren. Das Problem einer unbekannten Struktur von HP0565, wurde versucht durch Vorhersagen einer 2D-Topologie, abzudecken. An dieser Stelle ist die Qualität des vorhergesagten Modells und deren weiterführende topologische Charakterisierung durch spezifische Motive entscheidend. Aufgrund von unzureichenden Sequenzidentitäten zwischen Muster- und Zielstruktur konnte gegenwärtig kein ausreichend gutes 3D-Modell durch Informationen aus einer homologiebasierten Suche vorhergesagt werden. Deshalb muss ein korrektes Modellieren einer 3D-Struktur ohne Berücksichtigung von Anfangsbedingungen angeschlossen werden, was jedoch im Rahmen dieser Arbeit nicht möglich war. Eine Möglichkeit wäre eine Verbesserung der Modelle durch ein sogenanntes 2.5D-Modell. Indem gemessene Interaktionen der Helices und Wechselwirkungen mit der Membran berücksichtigt werden. Aus dem Experiment ist bekannt, welche Fragmente stabilisiert sind und somit wechselwirken können. Durch die Einbeziehung einer bekannten Membrandicke, wäre die Berechnung der Neigung von Helices möglich. Für eine zusätzliche Charakterisierung von bereits aufgestellten Stabilitätsbarrieren und erweiterten Sekundärstrukturelementen von HP0565 wird eine sequenzbasierte Energieprofilvorhersage verfolgt. Das Potential dieser Methode kann gegenwärtig den Informationsgehalt aus den Ebenen der Sequenz, der Topologievorhersage von HP0565 und beschriebenen Stabilitätskriterien vereinen.

Literaturverzeichnis

- [AMM04] ARAI, M. ; MITSUKE, H. ; M., Ikede: *ConPredII: a consensus prediction method for obtaining transmembrane topology models with high reliability*. Version: 2004. http://nar.oxfordjournals.org/cgi/reprint/32/suppl_2/W390, Abruf: 05.07.2010. – W390-W393, Nucleic Acids Research
- [BBAB02] BOECKMANN, B. ; BAIROCH, A. ; APWEILER, R. ; BLATTER, M.C: *The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003*. Version: 2002. <http://nar.oxfordjournals.org/cgi/reprint/31/1/365>, Abruf: 12.08.2010. – Vol. 31, no. 1, Nucleic Acids Research
- [BCF08] BOSSHART, P. ; CASAGRANDE, F. ; FREDERIX, P.: *High-throughput single-molecule force spectroscopy for membrane proteins*. 2008. – Nanotechnology, 19 384014 (15pp)
- [BWF99] BERMAN, H. ; WESTBROOK, J. ; FENG, Z.: *The Protein Data Bank*. Version: 1999. <http://nar.oxfordjournals.org/cgi/reprint/28/1/235>, Abruf: 12.08.2010. – Vol.28, no. 1, Nucleic Acids Research
- [CCJKJM06] CHIH-CHIEH, C. ; JENN-KANG, H. ; JINN-MOON, Y.: *(Ps)2: protein structure prediction server*. Version: 2006. http://nar.oxfordjournals.org/cgi/reprint/34/suppl_2/W152, Abruf: 09.07.2010. – W152-W157, Vol. 34, Nucleic Acids Research
- [Cle06] CLEMENT, J.: *Durin: Programm zur automatischen Analyse, Interpretation und Visualisierung von Daten kraftspektroskopischer Experimente an Membranproteinen*, Fachhochschule Lausitz, University of Applied sciences, Diplomarbeit, 2006
- [DCR05] DUNKER, A.K. ; CORTESE, M.S. ; ROMERO, P.: *Flexible nets-The role of intrinsic disorder in protein interaction networks*. Version: 2005. <http://onlinelibrary.wiley.com/doi/10.1111/j.1742-4658.2005.04948.x/pdf>, Abruf: 12.08.2010. – Issue 20, Volume 272, FEBS Journal
- [EM00] ENGEL, A. ; MULLER, D.J.: *Observing single biomolecules at work with the atomic force microscope*. 2000. – Nat. Struct. Bio. 7
- [Gae07] GAEDEKE, N.: *Biowissenschaftlich recherchieren: Über den Einsatz von Datenbanken und anderen Ressourcen der Bioinformatik (German Edition)*. 1. Birkhäuser Basel, 2007 <http://amazon.com/o/ASIN/>

- 3764385251/. – ISBN 9783764385255
- [Gal05] GALIP, K.: *Antigenerkennung während unterschiedlicher Stadien der Helicobacter pylori Infektion*. Version: 2005. <http://edoc.hu-berlin.de/dissertationen/karaali-galip-2005-06-24/HTML/front.html>, Abruf: 28.08.2009. Dissertation
- [Gmb] GMBH, Viro-Immun Labor-Diagnostika: *Helicobacter pylori*. <http://www.viro-immun.de>, Abruf: 31.08.2009
- [HBCM98] HIROKAWA, T. ; BOO-CHIEN, S. ; MITAKU, S.: *SOSUI: classification and secondary structure prediction system for membrane proteins*. Version: 1998. <http://bioinformatics.oxfordjournals.org/cgi/reprint/14/4/378?view=long&pmid=9632836>, Abruf: 06.07.2010. – Vol. 14, no.4, Bioinformatics Application
- [Hil03] HILDEBRAND, P.: *Zur Strukturvorhersage von Membranproteinen: Vergleich der Raumstrukturcharakteristika von Membranproteinen und globulären Proteinen*. Version: 2003. <http://edoc.hu-berlin.de/dissertationen/hildebrand-peter-2003-05-27/PDF/Hildebrand.pdf>, Abruf: 18.8.2009. Dissertation
- [HL03] HAESELER, A. ; LIEBERS, D.: *Molekulare Evolution*. 1. Fischer (Tb.), Frankfurt, 2003 <http://amazon.com/o/ASIN/3596153654/>. – ISBN 9783596153657
- [HL09] HUBERT, R. ; LETZEL, T.: *Der Experimentator: Proteinbiochemie/Proteomics*. 6. Spektrum Akademischer Verlag, 2009 <http://www.amazon.de/Experimentator-Proteinbiochemie-Proteomics-Hubert-Rehm/dp/3827423120>
- [JNe00] JANSHOFF, A. ; NEITZERT, F. ; ET.AL., Oberdörfer Y.: *Kraftspektroskopie an molekularen Systemen-Einzelmolekülspektroskopie an Polymeren und Biomolekülen*. 2000. – Angew. Chem., 112, 3346-3374
- [Jon07] JONES, D.T.: *Improving the accuracy of transmembrane protein topology prediction using evolutionary information*. Version: 2007. <http://bioinformatics.oxfordjournals.org/cgi/reprint/23/5/538>, Abruf: 05.07.2010. – Vol.23, no.5, Bioinformatics Original Paper
- [KG00] KANEHISA, M. ; GOTO, S.: *KEGG: Kyoto Encyclopedia of Genes and Genomes*. Version: 2000. <http://nar.oxfordjournals.org/cgi/reprint/28/1/27>, Abruf: 10.08.2010. – Vol. 28, no. 1, Nucleic Acids Research
- [KM08] KNOOP, V. ; MÜLLER, K.: *Gene und Stammbäume: Ein Handbuch zur molekularen Phylogenetik (German Edition)*. 2. neu bearb. und erw. Aufl. Spektrum Akademischer Verlag, 2008 <http://amazon.com/o/ASIN/3827419832/>. – ISBN 9783827419835

- [LBB07] LARKIN, M.A ; BLACKSHIELDS, G. ; BROWN, N.P.: *Clustal W and Clustal X version 2.0*. Version: 2007. <http://bioinformatics.oxfordjournals.org/cgi/reprint/23/21/2947>, Abruf: 12.08.2010. – Vol. 23, no. 21, BIOINFORMATICS
- [LCS03] LETUNIC, I. ; COPLEY, R. ; SCHMIDT, S.: *SMART 4.0: towards genomic data integration*. Version: 2003. <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC308822/pdf/gkh088.pdf>, Abruf: 12.08.2010. – Vol. 32, Nucleic Acids Research
- [Les02] LESK, A.M.: *Bioinformatik*. 1. Spektrum-Akademischer Vlg, 2002 <http://amazon.com/o/ASIN/3827413710/>. – ISBN 9783827413710
- [LMS08] LABUDDE, D. ; MARSICO, A. ; SCHROEDER, M.: *Stabilizing Regions in Membrane Proteins*. 2008. – Vol. 40, NIC Series
- [MA02] MULLER, D.J. ; ANDERSON, K.: *Biomolecular imaging using atomic force microscopy*. 2002. – Trends Biotechnol. 20, 45-49
- [Mül05] MÜLLER, F.: *Visualisierung von experimentellen Daten aus der Kraftspektroskopie unter der Verwendung der Topologie eines Membranproteins*. Dresden, Diplomarbeit, 2005
- [nan10] NANOMETIS: *nanometis* ® Software. <http://www.nanometis.com/services/>. Version: 2010
- [PEB04] PIEPER, U. ; ESWAR, N. ; BRABERG, H.: *MODBASE, a database of annotated comparative protein structure models, and associated resources*. Version: 2004. http://nar.oxfordjournals.org/cgi/reprint/32/suppl_1/D217, Abruf: 12.08.2010. – Database issue, Vol. 32, Nucleic Acids Research
- [Ser] SERVER, ExPASy P.: *Amino acid composition*. <http://www.expasy.ch/tools/protparam.html>, Abruf: 03.07.2010
- [SMR03] SELZER, P.M. ; MARHÖFER, R. ; ROHWER, A.: *Angewandte Bioinformatik: Eine Einführung (Springer-Lehrbuch) (German Edition)*. 1. Springer, 2003 <http://amazon.com/o/ASIN/354000758X/>
- [Thi04] THIES, C.: *Die Genotypen-Vielfalt in intergenischen Regionen des Bakteriums Helicobacter pylori- Aufbau von PCR-Assays zur Analyse der vac-A Promotor- Region und der Insertionsstelle der cag-Pathogenitätsinsel*. Version: 2004. http://www.freidok.uni-freiburg.de/volltexte/1930/pdf/aktuelle_Version_Doktorarbeit2005.pdf, Abruf: 28.08.2009. Dissertation
- [TI01] TUSNADY, G.E. ; I., Simon: *The HMMTOP transmembrane topology prediction server*. Version: 2001. [http://www.enzim.hu/~tusi/papers/Bioinformatics_17_849\(2001\).pdf](http://www.enzim.hu/~tusi/papers/Bioinformatics_17_849(2001).pdf), Abruf: 05.07.2010. – Vol.17, no.9, BIOINFORMATICS

- [Wer06] WERNER, J.: *Heterologe Expression von Antikörperfragmenten zur Kristallisation von Membranproteinen*. Version: 2006. http://deposit.d-nb.de/cgi-bin/dokserv?idn=988262681&dok_var=d1&dok_ext=pdf&filename=988262681.pdf, Abruf: 07.07.2010. Dissertation
- [Zha08] ZHANG, Y.: *I-TASSER Server for protein 3D structure prediction*. 2008. – BMC Bioinformatics

Selbständigkeitserklärung

Ich erkläre, dass ich die vorliegende Arbeit selbständig und nur unter Verwendung der angegebenen Literatur und Hilfsmittel angefertigt habe.

Mittweida, den 24.08.2010

Anne-Marie Pflugbeil